Addressing Uncertainty in Understanding Hydroclimate, Hydrology and Hydraulics Across Scales

by

Donghui Xu

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Civil Engineering)
in The University of Michigan
2020

Doctoral Committee:

Associate Professor Valeriy Y. Ivanov, Chair
Professor Avery Demond
Associate Professor Xianglei Huang
Assistant Professor Seymour M.J. Spence

Donghui Xu
donghui@umich.edu
ORCID iD: 0000-0002-2859-2664

For my family

ACKNOWLEDGEMENTS

This work was supported by Dow Sustainability Doctoral fellowship at the University of Michigan, Michigan Institute for Computational Discovery & Engineering Catalyst Grants, Google Earth Engine Research toward research project "Evapotranspiration of the Green Ocean Amazon", DOE Award DE-SC0011078 "Understanding the Response of Photosynthetic Metabolism in Tropical Forests to Seasonal Climate Variation", and NSF Grant EAR 1151443.

I wish to express my sincere appreciation to my advisor, Professor Valeriy Ivanov, who has the substance of a genius: he convincingly guided and encouraged me to be professional and do the right thing even when the road got tough. Without his persistent help, the goal of this dissertation would not have been realized. He always gives different options for projects for me to explore my research interests. Because of his flexibility, I had the opportunity to take many courses without the pressure of pushing forward research.

Besides my advisor, I would like to thank the rest of my dissertation committee members, Professor Avery Demond, Professor Xianglei Huang, Professor Seymour Spence, for their great support and invaluable advice on this dissertation. I am also grateful to Professor Nikolaos Katopodes, though I have to substitute you from the committee due to time conflict. Your courses and books sparked my interest in fluid mechanics.

I am also grateful to Professor Jingfeng Wang at the Georgia Institute of Technology, Dr. Khachik Sargsyan at the Sandia National Laboratories, and Professor

Simone Fatichi at the National University of Singapore. I am incredibly thankful and indebted to them for sharing expertise in evapotranspiration modeling, polynomial chaos expansion method, and stochastic downscaling techniques, respectively. The impact of their work on my own study is obvious throughout this dissertation.

My sincere thanks also go to Dr. Gautam Bisht, and Dr. Maoyi Huang, who provided me an opportunity to join Pacific Northwest National Lab as an intern, and who taught me parallel programming. Without their precious support, it would not be possible for me to develop a parallel overland flow model, which should be a long-term goal in the research.

I thank my fellow labmates in Michigan HYDROWIT group: Jongho Kim, Chase Dwelle, Elizabeth Agee, Frank Sedlar, Wenbo Zhou, Yuehan Wu, Weina Zhang, Liujing Zhang, I cannot begin to express my gratitude and appreciation for their friendship and academic support.

Most importantly, none of this could have happened without my family. Dad and Mom, thank you for providing me with unfailing support over these years. To my daughter, Anna, you are the most precious gift in my life. Accompany you to grow up is my luckiest thing. Finally, to my wife, Xinhang: your love and understanding helped me through the hard times. Without you believing in me, I would never have made it. It is time to celebrate; you earned this degree right along with me.

TABLE OF CONTENTS

LIST OF FIGURES

Figure

LIST OF TABLES

Table

LIST OF APPENDICES

Appendix

ABSTRACT

Human-induced climate change intensifies the hydrological cycle since air can hold more moisture at a higher temperature. Such effects can result in more frequent extreme weather conditions, such as floodings, droughts, and heatwaves. Outputs from the Global Climate Model are always used to study the impacts of climate change on the hydrological cycle. But the inferences suffer from significant uncertainties, which need to be addressed before any robust conclusion of climate change can be made. In this dissertation, a Bayesian weighted averaging (BWA) method is applied to investigate the change of peak annual runoff seasonality in the future period based on outputs from GCMs contributing to CMIP5. A high-quality daily runoff dataset is used in this BWA framework to reduce the bias of model projections. Based on the Bayesian inference, we identified a robust spatial pattern for the change of peak annual runoff timing, which is attributed to the shift of snowmelt and soil wetness seasonality caused by increased temperature.

Evapotranspiration (ET), representing a crucial component of the surface water, and energy balance, can be affected by climate change. The associated latent heat flux helps to control surface temperatures, which is related to the intensity and duration of heatwaves. However, the reliability of the estimation of ET at any scale remains challenging since observations of ET are sparse in space and time. Therefore, models are required to simulate ET at any scale using in situ or remote sensing observations. This dissertation applied a novel method based on the Maximum Entropy Production

(MEP) theory to estimate ET, which requires only net radiation, temperature, and specific humidity as inputs. Using site-level eddy flux data in the Amazon rainforest, the MEP method shows high skill at the hourly, daily, and monthly scales. The site-level MEP-based estimates outperform the estimates of the MODIS ET product, which is commonly used for large-scale assessments. We then applied MEP to project the change of ET in the future with GCMs forcings. The MEP-based projections are more robust than the GCM-based projections, implying the uncertainty of ET in climate models can be reduced by using this novel, parsimony algorithm.

It is clear that the hydroclimate is already changing, and changes will cascade into the watershed response. This dissertation presented a framework for quantifying the uncertainty of urban flooding simulation to resolve the uncertainty at the local scale. A physically rigorous hyper-resolution hydraulic and hydrologic model - tRIBS-OFM - is used here to advance numerical representation and understanding of interactions between flooding and the urban environment. Due to the steep computational cost and constraints of resolving the 2D Saint-Venant equations at very high resolutions, no effort has been made to address the uncertainty systematically. The uncertainty quantification remains challenging even the model is run in parallel with multiple cores. We approach this problem by training a surrogate model for tRIBS-OFM. Specifically, this surrogate model relies on polynomial chaos expansions, which creates a mapping of flooding outputs from the uncertain inputs. The surrogate model is very computationally inexpensive; therefore, the uncertainty in the inputs/parameters can be propagated to outputs efficiently through the surrogate model. Within this uncertainty quantification framework, we propose a real-time high-fidelity urban flooding forecasting framework, which is able to predict near-instantaneous quantities of interest (e.g., river discharge, inundation field) given a forecasted rainfall. For this study, we reproduce a flooding event in a catchment located in the city of Houston during the 2017 Harvey event and validate the performance of the uncertainty quantification

framework with streamflow, high watermarks, and inundation depth data.

CHAPTER I

Introduction

According to the Intergovernmental Panel on Climate Change report (IPCC, 2018), the earth is undergoing a period of rapid temperature increase caused by human activity. The global temperature rises about 1 °C compared to the pre-industries period, and it is projected to be worse in the future if we maintain the current CO_2 emission level. A warmer climate intensifies the hydrological cycle, which results in a more frequent occurrence of extreme events (e.g., flood, drought, heatwave). Therefore, it is imperative for us to understand how climate change affects hydroclimate, hydrology, and hydraulic across scales. For a larger scale, General Circular Models'(GCM) outputs are used directly to imply the impacts of climate change on the hydrological system (*Milly et al.*, 2005; *Chadwick et al.*, 2013; *Knutti and Sedláček*, 2013). However, the resolution of GCMs is too coarse to be directly used at local scales. The outputs need to be downscaled and then used as forcing in hydraulic/hydrologic/ecohydrology models to understand how the local system responds to climate change (*Kim et al.*, 2016; *Manning et al.*, 2009; *Kim and Ivanov*, 2015). During such analysis, uncertainties are inevitable, which should be addressed in some optimal fashion before any robust conclusions about climate change can be drawn.

1.1 Uncertainty Sources

1.1.1 Climate projections uncertainty

Coupled Model Intercomparison Project Phase 5 (CMIP5) provides outputs from about 50 GCMs developed by different institutes all over the world. As shown in Figure 1.1, significant discrepancies exist across the models when compared to observation at the same location. There is no 'perfect' climate model, therefore, it suggested climate projections should be inferred from multi-model ensemble (*Tebaldi and Knutti*, 2007). However, defining appropriate model weights for a multi-model ensemble remains challenging (*Christensen et al.*, 2010; *Knutti*, 2010; *Weigel et al.*, 2010). There are several solutions to assign weights for constructing a multi-model ensemble.

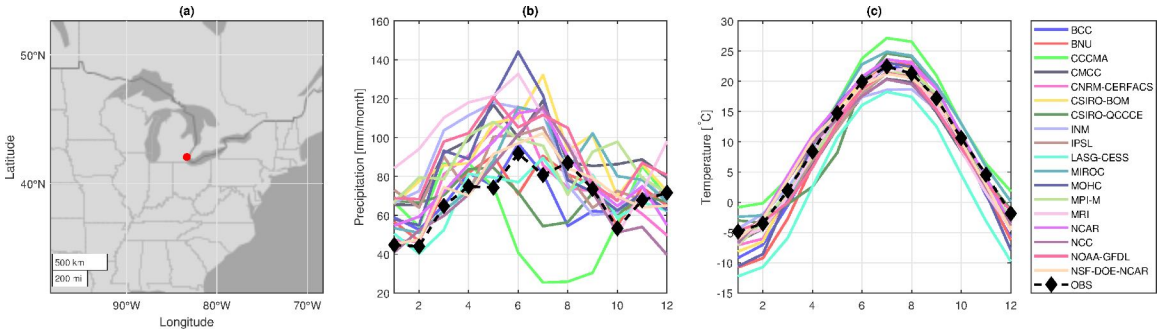

Figure 1.1: Precipitation seasonality (subplot b) and temperature seasonality (subplot c) of historical period (1961 - 1990) for the location of Detroit (red circle in subplot a) from each individual GCM participate CMIP5. The black dashed line with diamonds represent the observation for the same period from WebMET meteorological stations (`http://www.webmet.com/`).

The first approach is weighting each model equally (e.g., Simple Average, SA), and the ensemble uncertainty can be expressed with the variance. But this method is sensitive to outliers and is always associated with a large spread.

Bayesian formalism has shown the potential to be a sufficiently general machinery to combine multiple model estimates with varying degrees of uncertainty (*Fowler et al.*, 2007; *Fowler and Ekstrm*, 2009). The approach of *Raftery et al.* (2005) and

Gneiting et al. (2005) permits individual model bias correction and estimates model-specific weights as posterior probabilities relying on models performance in the training period (i.e., using historical observations). For example, this approach was applied to combine multiple Regional Climate Model (RCM) outputs in *Yang and Wang* (2011). However, it requires multiple replicates to use as training data, which may not always be available in hydrology and climate study.

Another method is the Bayesian Weighted Averaging (BWA) method of *Tebaldi et al.* (2005) that derived from the reliability ensemble average (REA) method (*Giorgi and Mearns*, 2002). Model weights are not defined here as posterior probabilities, but are assigned based on models performance in the control (i.e., accounting for model bias with respect to observations) and the future (i.e., accounting for the convergence of model results) periods (*Tebaldi et al.*, 2005; *Giorgi and Mearns*, 2002). The BWA methodology received significant attention because of its flexibility to use multimodel estimates with little subjectivism in terms of assumptions concerning the governing distributions. But recent studies argued this method narrows the uncertainty too much under some situations (*Smith et al.*, 2009; *Xu et al.*, 2018). To resolve this issue, a multivariate version of BWA that multiple regions are considered in one statistical model simultaneously was proposed by *Smith et al.* (2009).

A recent method of *Chandler* (2013) and *Leith and Chandler* (2010) differs from the above approaches by assuming that all climate models represent the same dynamical processes, though some may fail to reproduce the reality exactly. Specifically, the time series from different climate models are assumed to have a similar structure, sharing the same form of the statistical process. But parameters of the corresponding statistical descriptors can be different for different models. The associated uncertainty can be examined by deriving the distribution of these parameters.

1.1.2 Output downscaling

GCMs' resolutions are typically too coarse to be used directly at local scale (*Fowler et al.*, 2007; *Kim et al.*, 2016). Downscaling techniques aim to resolve these issues, and there is a need to further improve their skill in order to connect coarse-scale results to finer resolutions in terms of spatial and temporal scales. There are three common downscaling methods (*Fowler et al.*, 2007): dynamic downscaling (*Castro et al.*, 2005; *Jacob et al.*, 2014; *Maraun et al.*, 2010; *Wood et al.*, 2004), statistical downscaling (*Piani et al.*, 2010; *Schmidli et al.*, 2006; *Schoof and Pryor*, 2001; *Widmann et al.*, 2003; *Wilby et al.*, 1998; *Zorita and von Storch*, 1999), and stochastic downscaling (*Fatichi et al.*, 2013; *Kim et al.*, 2016; *Manning et al.*, 2009). They are outlined below.

In dynamic downscaling methods, GCM projections are used as the boundary conditions to drive models of finer spatial resolution, such as Regional Climate Model (RCM; *Christensen and Hewitson*, 2007). For example, the time-slice experiments of North American Regional Climate Change Assessment Program (NARCCAP) run simulations for the historical period using RCMs and boundary conditions reconstructed from observational data. Observations are subsequently perturbed using GCM climate trajectories to construct RCM boundary conditions for future periods (*Mearns et al.*, 2013). However, when forced with boundary conditions obtained from different GCM projections, RCM simulations can be significantly different (*Wang and Yang*, 2008; *Yang et al.*, 2012). This highlights that uncertainty originated from multiple trajectories determined by GCM outputs are not eliminated with RCM application and needs to be accounted for (*Olson et al.*, 2016; *Yang and Wang*, 2011). Few studies aimed to explicitly quantify the uncertainty using advanced methods. *Kang et al.* (2012) combined different RCMs from NARCCAP to investigate the uncertainty of their dynamical downscaling method. *Yang and Wang* (2011) used a Bayesian approach, concluding that the experiments yielding Bayesian ensemble mean were superior to RCM simulations driven by individual reanalysis products.

Still, even finer spatial scale RCM outputs are often too coarse for local impact studies such as those focusing on small and mid-sized watersheds, agricultural fields, or water infrastructure networks (*Chen et al.*, 2011). Another limitation of dynamic downscaling is that available outputs typically have a low temporal resolution (e.g., monthly) that do not capture extremes (*Maraun et al.*, 2010).

Statistical downscaling seeks to identify a relationship between GCM outputs and the corresponding observational data. The approach can be relatively simple and has received extensive attention (*Mamalakis et al.*, 2017; *Onyutha et al.*, 2016; *Jakob Themel et al.*, 2011). Compared to dynamic downscaling, it can reduce computational time significantly. Another advantage is that statistical downscaling is more flexible, it can be modified for any specific purpose, and observational data can be applied directly in its procedures (*Von Storch*, 2000). Although various statistical downscaling methods have been developed, they still pose multiple drawbacks. In essence, variability may be underestimated, while sensitivity to outliers can be too high, and many features of the projected climate variable are inherited from the historic data (*Diaz-Nieto and Wilby*, 2005; *Fowler et al.*, 2007).

Stochastic downscaling approaches offer an attractive alternative to address the deficiencies of the above methods (*Burlando and Rosso*, 2002; *Fowler et al.*, 2007; *Maraun et al.*, 2010; *Semenov and Stratonovitch*, 2010). In these methods, the variability of climatic time series is fully explored, as compared to statistical downscaling. As a specific example, an hourly weather generator that can reproduce stochastic variations in climate variables (*Fatichi et al.*, 2011; *Ivanov et al.*, 2007), coupled with a Bayesian Weighted Averaging (BWA) method (*Tebaldi et al.*, 2005), was developed to downscale climate projections from native GCM resolutions to station-scale (*Fatichi et al.*, 2013; *Kim et al.*, 2016). Weather generators have been applied to locations with different climate characteristics showing highly satisfactory performance (*Fatichi et al.*, 2011, 2013, 2016; *Fowler et al.*, 2007; *Kim et al.*, 2016; *Manning et al.*, 2009;

Peleg et al., 2017). Specifically, observational data are required to estimate a set of parameters that allow the generator to mimic historical climate conditions. Factors of change (FOCs) (*Fatichi et al.*, 2011) are subsequently applied to climate statistics representing historical observations to yield statistics representative of the future period, whose philosophy is similar to the 'delta change factor' and 'change factor' presented in *Akhtar et al.* (2008); *Anandhi et al.* (2011); *Hay et al.* (2000); *Kilsby et al.* (2007); *Fowler et al.* (2007); *Fowler and Ekstrm* (2009). Two kinds of FOC are used in the stochastic downscaling framework, additive factor of change is the difference between a GCM statistic representing the future period (FUT) and the control period (CTL); it is typically used for statistics of air temperature. Product factor of change is the ratio between GCM statistics for the future and control periods, and it is commonly used for statistics of precipitation. The product factor of change may also be applied to other variables such as vapor pressure, incoming shortwave radiation, cloudiness, wind speed, and atmospheric pressure required by weather generator. The assumption behind such an approach is that GCMs produce reasonable estimates of the change for various climate statistics, even though not necessarily accurate absolute magnitudes (*Anandhi et al.*, 2011; *Fowler et al.*, 2007). Additionally, FOCs are calculated from multiple outputs of GCMs (or RCMs) using BWA (*Fatichi et al.*, 2011; *Fowler and Ekstrm*, 2009; *Kim et al.*, 2016) or other techniques (*Bishop and Abramowitz*, 2013; *Haughton et al.*, 2014, 2015; *Semenov and Stratonovitch*, 2010), to reduce the uncertainty associated with the multi-model ensemble spread (*Tebaldi and Knutti*, 2007).

1.1.3 Hydrologic model uncertainty

Due to the complexity of the hydrologic system, it is impossible to reproduce any corresponding process exactly in a model. Each model simplifies the physical process to some extent, to guarantee a tractable solution. Consequently, these assumptions

introduce uncertainty to the predictions, which can only be reduced by improving the model with better knowledge. Both the data used as input to the model and the data used to validate the model always contain uncertainty caused by sampling and measurement errors. The uncertainties and limited length of data can lead to uncertain model parameters estimates during calibration. Further, one should expect more uncertainty in the outputs when extending such a calibrated model to an ungauged watershed with specified parameters or empirical values, which is common in hydrology studies. Therefore, it is essential to account for uncertainty when one predicts with a model.

Numerous approaches have been developed to quantify the predictive uncertainty of hydrologic responses (*Renard et al.*, 2010; *Montanari and Koutsoyiannis*, 2012; *Beven*, 1993), but this remains a challenge. For example, finer spatial resolutions of relevant data sets on topography, land cover, soil type, building layout, and civil infrastructure describe the heterogeneity of the real world and have the potential to advance the knowledge on hydrological/hydraulic process in natural and urban areas, but also come at an extreme computational cost. In the traditional uncertainty quantification method, one needs to draw samples from the priors of model parameters and inputs to drive the model, for example, through Monte Carlo simulation. Consequently, thousands of simulations have to be run to obtain the uncertainty in the outputs because of the high dimensionality of the parameter space and heterogeneity of input fields. The traditional method for uncertainty quantification is not always feasible for the high-fidelity model, which is computationally expensive.

Recent work of *Dwelle et al.* (2019) applied the framework of *Najm* (2009) and *Sargsyan et al.* (2014) to quantify uncertainty of a physically hydrologic model, tRIBS-VEGGIE (*Ivanov et al.*, 2008, 2010), which is computationally expensive. Specifically, Polynomial Chaos Expansion (PCE) is used to build surrogate, reduced-order representation of a physical model. A surrogate model is generally used to simulate the

behavior of a more computationally complex model. The purpose of using a surrogate model is that it is computationally inexpensive compared to the original model, and it can be rigorously sampled for uncertainty propagation, parameter inference, or sensitivity analysis. A full analysis of the uncertainty for any outputs of interest is carried by running a Monte Carlo simulation with the surrogate model. Due to the efficiency of the PCE surrogate, thousands of simulations can be finished in a reasonable time. There are multiple frameworks that fall into the class of surrogate models, e.g.: Gaussian process (GP) models (*Rasmussen*, 2004; *Kennedy and O'Hagan*, 2000), artificial neural networks (*Ripley*, 1996), support vector machines (*Abe*, 2010), and polynomial chaos expansions (PCEs) (*Xiu and Karniadakis*, 2002). GP models and PCEs are mainly used in statistics and engineering (often computational fluid dynamics), whereas neural networks and support vector machines were more used in data-intensive applications.

1.2 Research Scope

This dissertation aims to address uncertainties in hydrological applications across different scales, for example, assessing the impacts of climate change on hydrological variables at coarse scale and predict the response of a watershed to an extreme event at local scale. In Chapter II, the stochastic downscaling framework of *Fatichi et al.* (2013) and *Kim et al.* (2016) is revisited. It is applied to three locations of the US with different climate characteristics to investigate the role of observation in this framework. The first step of the stochastic downscaling framework is extracting climate signals in the future from climate models with Bayesian Weighting Averaging (BWA) method *Tebaldi et al.* (2005). Observations are generally thought to be important for correcting bias in BWA multimodel ensemble. However, we found the inclusion of observation in the BWA procedure (*Tebaldi et al.*, 2005) can introduce unpredictable impacts, such as unrealistic narrow uncertainty. A modified BWA method

is proposed to use in the stochastic downscaling framework, differences between the modified method and the original one are discussed. Further, the comparison implies that the modified BWA produces more reasonable uncertainty bound for Factor of Changes.

We apply a multivariate BWA framework (*Smith et al.*, 2009) in Chapter III to infer the change of peak annual runoff timing with daily runoff projections from 10 CMIP5 GCMs. This framework is adapted to the time variable due to its circular nature. Based on the Bayesian inference, we identified a clear spatial change pattern for the peak annual runoff timing in the future caused by global warming. Specifically, the areas where runoff is dominated by snowmelt will experience an earlier peak runoff. For the areas that the peak annual runoff timing is projected to delay, we found that soil moisture plays an important role.

Chapter IV validates a novel evapotranspiration model based on the Maximum Entropy Production (MEP) theory with nine observation sites in the Amazon rainforest. While MEP requires only net radiation, temperature, and specific humidity as inputs, it shows high skill in reproducing ET in highly biodiverse regions. The uncertainty of MEP based ET is less than that of the classical ET algorithm, which requires inputs such as wind, roughness, resistance parameters, etc. Those inputs or parameters are hard to obtain on the surface and can carry high uncertainty, especially in high diversity rainforest. In light of the parsimony and good skill of the MEP model, we apply it to project the change of ET at the US continent scale with inputs from GCMs.

In Chapter V, Polynomial Chaos Expansion (PCE) framework is coupled with a high-fidelity urban flooding model (e.g., tRIBS-OFM) for uncertainty quantification. Simulations from tRIBS-OFM are used to train the PCE based surrogate model, which creates a mapping of flooding outputs from the uncertain inputs. The surrogate model is computationally inexpensive due to its polynomial forms, making it

feasible for real-time flood forecasting with uncertainty warranted. To demonstrate the performance of this framework, we reproduce a flooding event in a catchment located in the greater Houston area during the 2017 Harvey event and validate its performance of forecast with USGS streamflow and inundation depth data.

The last chapter summarizes the dissertation and numerates the assumptions and limitations, followed by the future direction of the research. Specifically, Chapter II, III, and IV rely on the outputs from CMIP5 to infer the impacts of climate change on horological variables. The models of new versions, Coupled Model Intercomparison Project Phase 6 (CMIP6), are available to download and use recently. A comparison between CMIP5 and CMIP6 is important to understand if the skill of the newer version is improved, which provides useful information for the modeler. The increasing abundance and quality of data sources support the possibility of future regional and global heat flux assessments with the Maximum Entropy Production (MEP) theory. Geostationary Operational Environmental Satellite provides high-resolution observations of net radiation and temperature, which can be used as inputs for the Maximum Entropy Production method. The high-resolution MEP-based evapotranspiration estimates over the Amazon basin can be used to examine the response of plants to droughts, for example, the 2015-2016 mega drought. Lastly, the uncertainty quantification framework is not without its challenges. Although there is a reduction in the number of computational resources required during the real-time forecast, high-performance computing (HPC) resources are still required to train the surrogates. Additionally, the number of required training simulations grows rapidly when the number of uncertain inputs increases. Other surrogate techniques rely on less training simulations should be considered as an alternative in future works.

CHAPTER II

On the Use of Observations in Assessment of Multi-model Climate Ensemble

2.1 Introduction

Climate change will affect global and local temperatures and the distribution and amount of precipitation, with direct consequences for regional hydrology and water resources in many parts of the world (*Allan et al.*, 2014; *Hanson et al.*, 2012; *Kim et al.*, 2016; *Mahlstein et al.*, 2012; *Nunes et al.*, 2013; *Safeeq and Fares*, 2012). It is thus vital to quantify characteristics of the change and the corresponding uncertainty. A substantial amount of recent research has relied on climate projections obtained with General Circulation Models (GCMs) that contributed to the fifth phase of the Coupled Model Intercomparison Project (CMIP5; *Taylor et al.*, 2012b) or CMIP3 (*Meehl et al.*, 2005). Such modeling results typically carry biases that must be reduced in some optimal fashion before any robust conclusions of climate change can be drawn (*Giorgi and Mearns*, 2002; *Knutti*, 2010; *Knutti et al.*, 2017; *RISNEN*, 2007). Additionally, outputs from GCM are too coarse to be applied directly at local scales.

In this chapter, a stochastic downscale method is used to address the resolution issue. Specifically, climate change information is extracted using the BWA approach of *Tebaldi et al.* (2005) in the form of factors of change (FOCs), which are used to perturb

the observation series and obtain parameters representative of future climate for an hourly weather generator as described below. Previous studies using the approach with GCM or RCM ensembles obtained FOCs and then applied these factors to the same observations that had been used in BWA to infer the climate signal (*Fatichi et al.*, 2011, 2013; *Fowler et al.*, 2007; *Kim et al.*, 2016; *Manning et al.*, 2009). Other techniques relying on a multimodel ensemble to infer climate change information have also been evaluated and discussed (*Collins et al.*, 2012; *Haughton et al.*, 2015; *Knutti*, 2010; *Leith and Chandler*, 2010). While these studies carried out comprehensive analyses, a focused discussion of the role of observations in the context of Bayesian averaging of model result ensemble has not yet taken place. This appears to be particularly important for such a vital climate variable as precipitation.

FOCs can be extracted from GCMs with or without relying on observation to correct the bias. The default methodology of *Tebaldi et al.* (2005) accounts for biases (i.e., uses observations in the estimation of the PDFs of factors of change), but we find that this creates vulnerability for cases when a substantial disagreement exists between model outputs and observed data. We also contend that these caveats are not rare cases but typical for CMIP5 precipitation projections at the national scale. Although observations are considered to be crucial for correcting biases in GCM outputs (*Tyralis and Koutsoyiannis*, 2017), one can surmise that GCMs provide reasonable climate signals without reproducing observations accurately (*Anandhi et al.*, 2011; *Fowler et al.*, 2007). Furthermore, using observations twice in the stochastic downscaling framework is not advisable. The above issues suggest that observations should be excluded in BWA when FOCs are computed; however, historical observations are necessary when FOCs are applied to calculate the characteristics of the projected climate.

We propose a modified version of the BWA approach that isolates the stage of identification of projection information based on outputs of climate model ensemble

(historic vs. future periods), from the step of observation-based weighting. Observations are used in the last phase of the stochastic downscaling framework by perturbed by the inference of the climate change signal. Such an approach is consistent with the philosophy that is already used in dynamic downscaling (e.g., in NARCCAP efforts): variation between the current and projected future climates is evaluated from the model outputs and only then it is applied to observations to yield a projection of future conditions (*Mearns et al.*, 2013). Additionally, two alternative methods, ensemble "simple average" and Bayesian averaging with equal weights, are introduced as a reference for the revised BWA method. It should be noted that the proposed BWA doesn't represent a generalized Bayesian weighting method, but rather a method that is more suitable in frameworks of stochastic downscaling.

2.2 Ensemble averaging

A detailed stepwise procedure of the stochastic downscaling method used in this chapter is presented in *Kim et al.* (2016) and *Fatichi et al.* (2011). Only a brief description of the original BWA methodology, related parameters, and modifications are shown in the following.

2.2.1 Bayesian Weighted Averaging

The BWA approach of *Tebaldi et al.* (2005) has grown in popularity as a sufficiently general tool with minimum subjective assumptions to assess uncertainty in climate change from multiple model projections. It relies on independent GCM outputs, which is important for multi-model ensemble prediction (*Abramowitz*, 2010), and in this chapter realizations from 18 GCMs developed in different institutions are downloaded from the CMIP5 database (http://pcmdi9.llnl.gov/, see Table 2.1). We picked one model from one institution to reduce the dependence among models. Mean temperature and daily precipitation statistics are calculated from GCM realizations

for each month of the year. Posterior distributions for the control (CTL) and future (FUT) scenarios are generated by BWA, with the weight estimated based on the performance of the models in reproducing historical observations as well as inter-model agreement in the future period. These criteria are referred to as the model 'bias' and 'convergence', respectively. A model with a smaller bias in the CTL period and a better agreement with other model estimates in the FUT period is weighted more heavily in the estimation of the posterior distributions of a variable of interest (such as mean air temperature or mean precipitation) for both periods.

The approach of *Tebaldi et al.* (2005) is described briefly in the following. Specifically, model outputs are assumed to follow Normal distribution.

$$\begin{aligned} X_i &= \mu + \eta_i \\ Y_i &= \nu + \beta_x(X_i - \mu) + \xi_i/\sqrt{\theta} \end{aligned} \tag{2.1}$$

where $\eta_i \sim N(0, \lambda_i^{-1})$, and $\xi_i \sim N(0, \lambda_i^{-1})$. In Eq (2.1), X_i and Y_i represent ith model output of climate variables for the control and future periods respectively, and parameters μ and ν are the corresponding true values, which are common for all GCMs.

Model weight is not defined explicitly in the BWA framework but is related to a 'precision parameter' $\lambda_i (i = 1 \ldots 18)$ and λ_0, where the subscript i represents the ith GCM model and λ_0 is related to the weight for observational data. The θ is the parameter that allows GCMs to have different precisions when simulating future climate conditions, as compared to the historical period; *Tebaldi et al.* (2005) uses β_x as a regression parameter to introduce correlation between the model outputs of past

and future. The prior distribution is selected for each parameter:

$$\begin{aligned} \mu, \nu, \beta_x &\sim U(-\infty, \infty) \\ \lambda_i &\sim Ga(a, b) \\ \theta &\sim Ga(c, d) \end{aligned} \tag{2.2}$$

Uniform distribution and Gamma distribution (Ga(a,b)) with $a = b = c = d = 0.01$ are adopted to guarantee the priors are uninformative. Then, we can have posterior distributions from the Bayes theorem:

$$Posterior \propto Prior \times Likelihood \tag{2.3}$$

Although the analytical forms of the joint posteriors for the climate quantities of interest are unknown, closed-forms of each marginal posterior distribution are derived in the appendix of *Tebaldi et al.* (2005). Markov Chain Monte Carlo (MCMC) process is used to estimate the posterior distributions. Specifically, 25,000 iterations are used as "burn-in" period here, and sampling of the posterior distribution is saved every 50 iterations to minimize correlation between successive samples. Finally, 1000 samples are used to construct the posterior distribution of each parameter, which results in 75,000 iterations in total. Parameter λ_i is sampled from Gamma distribution with expected value of form in Eq (2.4), which allows its estimation including the uncertainty (*Tebaldi et al.*, 2005). Parameter λ_0 is constant value as a measure of natural variability for the variable, which is determined from historic data (*Fatichi et al.*, 2011).

$$E(\lambda_i) = \frac{a+1}{b + \frac{1}{2}(X_i - \tilde{\mu})^2 + \frac{1}{2}(Y_i - \tilde{\nu} - \beta_x(X_i - \tilde{\mu}))^2} \tag{2.4}$$

Parameters a and b are set to be 0.01 to ensure the prior distributions for λ_i are

diffuse/non-informative. In above equation, $\tilde{\mu}$ and $\tilde{\nu}$ (see Eq (2.5)) are the expected values of μ and ν. Consequently, the terms $(X_i - \tilde{\mu})^2$ and $(Y_i - \tilde{\nu})^2$ correspond to the 'bias' and 'convergence' criteria, respectively. In the last step, FOC distributions are constructed using the BWA-computed posterior distributions for the control and future periods.

$$\begin{aligned} \tilde{\mu} &= \frac{\sum_{i=1}^{N} \lambda_i X_i - \theta\beta_x \sum_{i=1}^{N} \lambda_i (Y_i - \nu - \beta_x X_i) + \lambda_0 X_0}{\sum_{i=1}^{N} \lambda_i + \theta\beta_x^2 \lambda_i + \lambda_0} \\ \tilde{\nu} &= \frac{\sum_{i=1}^{N} \lambda_i (Y_i - \beta_x (X_i - \mu)}{\sum_{i=1}^{N} \lambda_i} \end{aligned} \tag{2.5}$$

Table 2.1: A list of CMIP5 models used in this chapter

No.	Institution	Model name	Resolutions	References
1	BCC	bcc-csm1-1	128 × 64	*Wu* (2012)
2	BNU	BNU-ESM	128 × 64	*Wei et al.* (2012)
3	CCCMA	CanESM2	128 × 64	*Chylek et al.* (2011)
4	CMCC	CMCC-CM	480 × 240	*Scoccimarro et al.* (2011)
5	CNRM_CERFACS	CNRM-CM5	256 × 128	*Voldoire et al.* (2013)
6	CSIRO_BOM	ACCESS1-0	192 × 145	*Bi et al.* (2013)
7	CSIRO_QCCCE	CSIRO-Mk3-6-0	192 × 96	*Jeffrey et al.* (2013)
8	INM	Inmcm4	180 × 120	*Volodin et al.* (2010)
9	IPSL	IPSL-CM5A-LR	96 × 96	*Dufresne et al.* (2013)
10	LASG_CESS	FGOALS-g2	128 × 60	*Li et al.* (2013)
11	MIROC	MIROC5	256 × 128	*Watanabe et al.* (2010)
12	MOHC	HadGEM2-ES	192 × 145	*Jones et al.* (2011)
13	MPI_M	MPI-ESM-MR	192 × 96	*Giorgetta et al.* (2013)
14	MRI	MRI-CGCM3	320 × 160	*Yukimoto et al.* (2012)
15	NCAR	CCSM4	288 × 192	*Gent et al.* (2011)
16	NCC	NorESM1-M	144 × 96	*Bentsen et al.* (2013)
17	NOAA_GFDL	GFDL-ESM2G	144 × 96	*Dunne et al.* (2012)
18	NSF_DOE_NCAR	CESM1-CAM5	288 × 192	*Meehl et al.* (2013)

The purpose of this study is to investigate the effect of observations in the identification of climate change signals from multiple GCM outputs in the context of Bayesian averaging inference. In this chapter, we propose to remove the observation-based weight (i.e., λ_0) in the formulation of the original BWA framework and introduce observational information after FOCs are estimated using GCM outputs solely in

the stochastic downscaling framework. The formulation for computing the posterior expected value $\tilde{\nu}$ (future period) remains identical to the expression in *Tebaldi et al.* (2005) but the estimation of the expected value corresponding to the historic period $\tilde{\mu}$ now follows Eq (2.6). The 'bias' criterion is removed from the weighting scheme that impacts the estimation of the posterior distributions, and therefore model weights stem only from the model output convergence at the control and future periods. As a result, $\tilde{\mu}$ has a similar formulation as $\tilde{\nu}$, with the term $(X_i - \tilde{\nu})^2$ representing the deviation of the *ith* model projection from the overall 'convergence' with respect to the model ensemble at the control period.

$$\tilde{\mu} = \frac{\sum_{i=1}^{N} \lambda_i X_i - \theta\beta_x \sum_{i=1}^{N} \lambda_i (Y_i - \nu - \beta_x X_i)}{\sum_{i=1}^{N} \lambda_i + \theta\beta_x^2 \lambda_i} \tag{2.6}$$

Defining appropriate model weights for a multi-model ensemble remains challenging (*Christensen et al.*, 2010; *Knutti*, 2010; *Weigel et al.*, 2010). To provide more context for differences among various weighting schemes, ensemble Simple Averaging (SA) and Simple Averaging in the Bayesian framework (BSA) are also used here for a comparison. SA represents the arithmetic average of all GCM projections in the model ensemble. BSA assigns equal weights to model outputs using the BWA framework - this is in contrast to the weights that depend on the individual model performance (*Fatichi et al.*, 2016). Eq (2.7) is used to determine the precision parameter for each model in the BSA context.

$$\lambda_{BSA} = \frac{1}{0.5 \left(\frac{\sum_{i=1} N(X_i - \bar{X})^2}{N} + \frac{\sum_{i=1} N(Y_i - \bar{Y})^2}{N} \right)} \tag{2.7}$$

Equal weights in the BWA framework further imply that the precision parameter is the same for all models in the GCM ensemble. The denominator in Eq (2.7) is the average of estimates of model output variances for the control and future periods,

indicating that the uncertainty of the posterior distributions is related to the overall variability in GCM outputs. In an analogy with Eq (2.4), outputs from both periods are used to estimate the weights.

2.2.2 Data

Data from three exemplary locations are used in this chapter: Fresno, CA (36.780°N, 119.719°W), Miami, FL (25.917°N, 80.283°W), and Flint, MI (42.969°N, 83.752°W). These locations represent different climatic characteristics and have detailed hourly meteorology data, which are required for parameterizing weather generator. The location of Fresno has Mediterranean characteristics with warm, wet winters and hot, dry summers; Miami has a warm climate and high precipitation throughout the year; the location of Flint represents typical humid, temperate, continental climate (*Peel et al.*, 2007). RCP45 and RCP85 are chosen as the emission scenarios due to the availability of all 18 selected GCMs. Two future periods are simulated: 2041-2070 and 2071-2100 to represent the middle and end of the century. Control (CTL) period is defined as 1961-1990, and observational data for precipitation and temperature for the control period are downloaded from Meteorological Resource Center (`www.webmet.com`).

2.3 Results

2.3.1 Original BWA results of estimating FOCs for precipitation

The caveats of including observations in the estimation of FOC distribution are illustrated in Figure 2.1a. It shows that all models fail to reproduce the observation. As a result, the posterior distribution for the CTL estimated using the approach of *Tebaldi et al.* (2005) has most of its mass shifted away from the cluster of model outputs towards the observation. This is because the majority of the models have larger biases, and their weights are thus penalized. The outputs from two models

(#3 and #7) that are the closest to the observation receive higher weights, even though they also do not reproduce the observation well. As seen, the shift of the PDF also propagates into the posterior distribution for future precipitation (Figure 2.1b). This is questionable since the implication is that too much confidence (about 1 or 2 orders of magnitude higher, see Figure 2.1d) is placed in the two models, while the information contained in the outputs from other models is mostly ignored.

The situation of a substantial disagreement between the model reconstruction of the historical past and actual observations is not an exception. As Figure 2.2 clearly shows, the GCM precipitation ensemble fails to capture station-level observations most of the time: one can see a small slope of the empirical cumulative distribution function in the region of 50% (Figure 2.2a). Specifically, models tend to overestimate precipitation in the U.S. Northeast, while underestimating it in the U.S. Southeast. As a result, the traditional BWA approach tends to over-emphasize the importance of few models whose outputs "happen" to be closest to the observations. While there could be a favorable physical explanation of the better match of these model results with observations, placement of confidence occurs without consideration of that. Furthermore, the value of the information contained in other model outputs is downplayed.

2.3.2 Performance of revised BWA

We focus on mean monthly precipitation and temperature because these are some of the most significant hydro-meteorological variables. Figure 2.3 illustrates the uncertainty of results for CTL and FUT periods using the 5 (lower bound) and 95 (upper bound) percentiles of the BWA posterior distributions for the original and modified methods. The difference is most pronounced for the distributions of the CTL and FUT scenarios, while it is less pronounced for the FOC distribution since the effect of observation between the two periods is partially canceled out by computing the ratio.

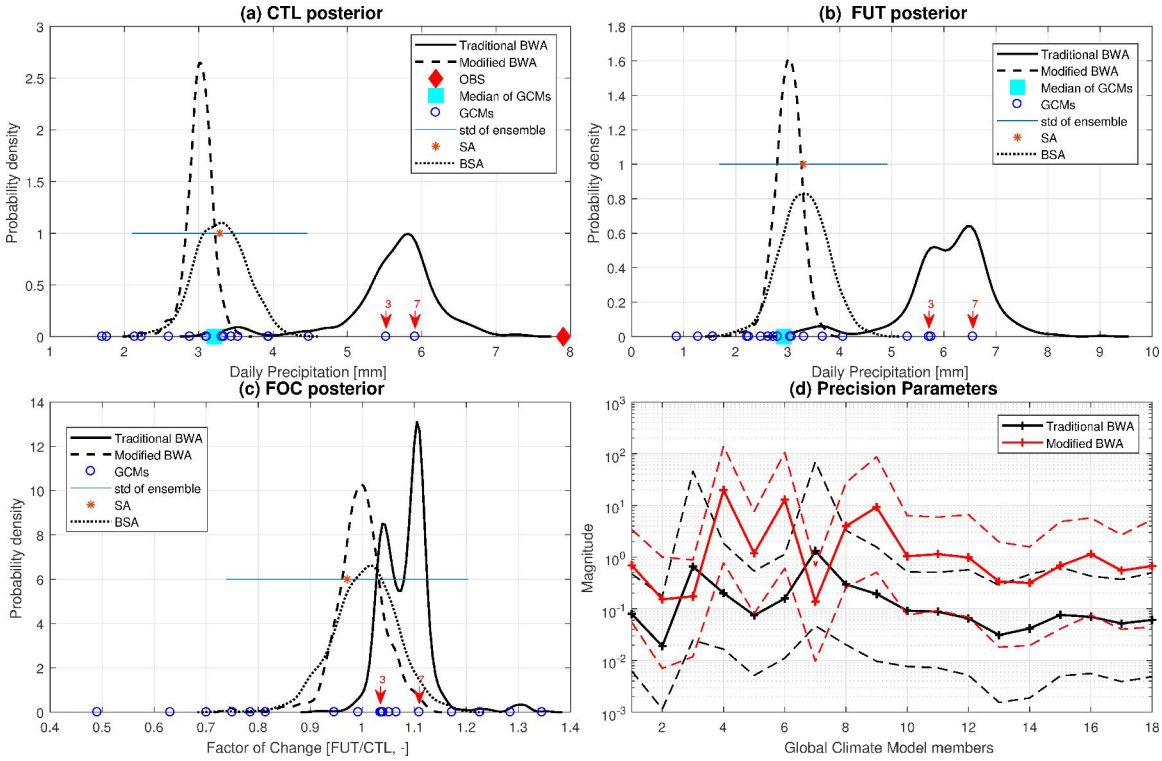

Figure 2.1: Posterior distributions (solid and dashed lines), individual GCM results (circle symbols), and observations (red diamond symbol) for mean June precipitation for the location of Miami corresponding to **(a)** control period (1961-1990); **(b)** future period (2041-2070, CMIP5 RCP45 scenario); and **(c)** the PDF of factor of change [FUT/CTL]. Subplot **(d)** shows the median (solid lines) and the 5th-95th percentiles (lower and upper dashed lines) for GCM 'precision parameters' of *Tebaldi et al.* (2005) (λ_i) based on 1000 MCMC samples. In subplots **(a)-(c)**, '3' and '7' represent the model number in Table 2.1. Traditionall BWA' indicates Bayesian Weighted Averaging method in *Tebaldi et al.* (2005); 'Modified BWA' indicates the method presented in this chapter; 'SA' is simpe averaging of ensemble of model prejections; and 'BSA' indicates simple averaging, where constant, pre-determined model weights are used in Bayesian context. 'OBS' refers to the observation.

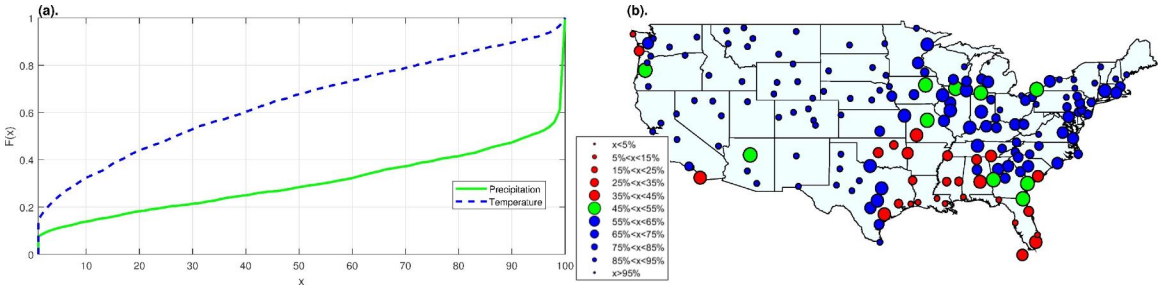

Figure 2.2: The impact of observations on the shift of the CTL posterior distribution of mean monthly precipitation (subplot **(a), (b)**) and temperature (subplot **(a)** only). The posterior distribution is estimated using the approach of *Tebaldi et al.* (2005). Observations are point measurements from 171 stations obtained from the Meteorological Resource Center archive (`www.webmet.com`) for the continental U.S.: 171 (stations) $\times$ 12 (months) = 2502 station-months, 30-year averages are used. In subplot **(a)**, the X-axis represents the percentile of the estimated CTL posterior distribution (precipitation or temperature) corresponding to the median of individual model outputs. For example, 50% implies that the center of mass of the posterior distribution approximately coincides with the center of mass of the GCM output ensemble implying that model outputs tend to agree well with observations; the observation is larger for lower percentile values (such as illustrated in Figure 2.1a) or smaller for higher percentiles. For example, 0% and 100% imply that the observation is outside of the model output ensemble, and therefore, the posterior distribution is shifted away from the median of the model outputs, effectively ignoring information content in GCM simulations. The Y-axis is the empirical cumulative distribution function estimated using 2051 values. The slope of the green line is smaller than that of the blue line, illustrating the level of agreement for precipitation is smaller than for temperature. Subplot **(b)** shows the spatial distribution of 171 stations. The symbol size (as shown in the model ensemble outputs, e.g., values around 50% is largest and around 0% and 100% is the smallest. The symbol color indicates the direction of the deviation: green means 50%, blue means higher percentile (models overestimate), and red means lower percentile (models underestimate)

However, for a few months when GCMs do not well reproduce observations (e.g., observations for the month of June-August are at the margin of the model ensemble), the differences between the original and modified methods in terms of the median of FOCs are quite appreciable (e.g., 9%, 14% and 16% for respective months). These differences are quite large because product FOC will be multiplied by observation in the subsequent steps. Figure 2.1 explicitly shows a comparison of two methods for the month of June. As Figure 2.1d indicates, the majority of the model precision parameters λ_i for the original method are lower than that of the revised approach, except for models #3 and #7. These two models are considered to be more 'reliable' (resulting in higher weights in the estimation of the posterior distribution) since their outputs are closer to the observational values, even though they are outliers in terms of model output convergence. In Figure 2.1c, it can also be seen that the FOC distribution inferred using the original BWA shows bi-modality, which results from the 'conflict' of the two criteria, i.e., the 'bias' and 'convergence', in the weight assignment. The reason is that model #3 and #7 gain higher weight than other models, but their future projections are not consistent with the rest models (i.e., other models exhibit larger convergence). The bi-modal posterior distribution is also a questionable outcome since the assumption of the posterior distribution is its normality (*Tebaldi et al.*, 2005); and a specified percentile (i.e., 50%) of the bi-modal distribution cannot be used to infer the corresponding properties at maximum likelihood. Conversely, after the weight of observations has been removed from this procedure, most of the 18 models receive larger weights, and the precision of model 'outliers' (i.e., models #3 and #7) drops. Therefore, the revised BWA leads to a conceptually consistent posterior distribution of FOC, as only models with higher output convergence get the higher weights.

For the case when model outputs agree with the observation, a comparison between two methods is quite similar and model precision parameters are essentially the

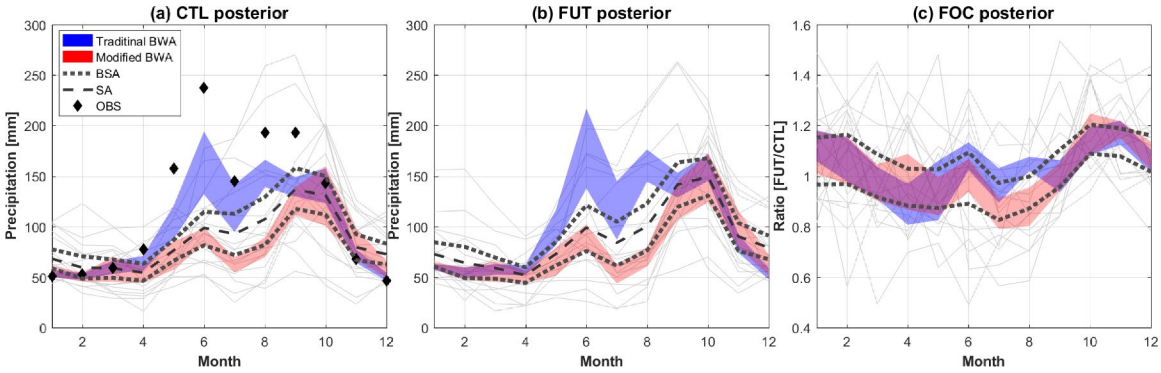

Figure 2.3: The 5th-95th percentile uncertainty bounds (blue and red) of the posterior distributions of mean precipitation for the location of Miami corresponding to **(a)** control period (1961-1990); **(b)** future period (2041-2070, RCP45 scenario); and **(c)** the PDF of FOC [FUT/CTL]. The light grey solid lines correspond to individual GCM monthly precipitation seasonalities. The two dotted dark grey lines are the 5th-95th percentile uncertainty bounds for the BSA method. 'OBS' refers to the observed data.

same (see Figure 2.4). Therefore, the revision impact is minimal for such cases and the proposed method is asymptotically consistent with that of *Tebaldi et al.* (2005).

For the location of Fresno that exhibits several months with deficient precipitation, the results are qualitatively consistent with the above considerations (Figure 2.5). For the dry summer period, the product FOC is very sensitive to the amount of the projected change. A small change of precipitation in the future results in a very large FOC, because the absolute value of the CTL period can be tiny, or even zero (i.e., resulting in infinite product FOC). Since all GCMs overestimate precipitation, the posterior distribution for the control period based on the traditional BEA shifted towards the observation, which is outside of the model ensemble. The effect of observation also propagates to future posterior, i.e., the mean of the posterior is much lower than most of the individual model projections.

As before, even though a large difference exists between the traditional and the revised BWA for the control and future scenarios, the posterior FOC distributions are similar for most of the months. However, for the driest month (June), the posterior distributions of FTU and FOC contain negative values, which are physically

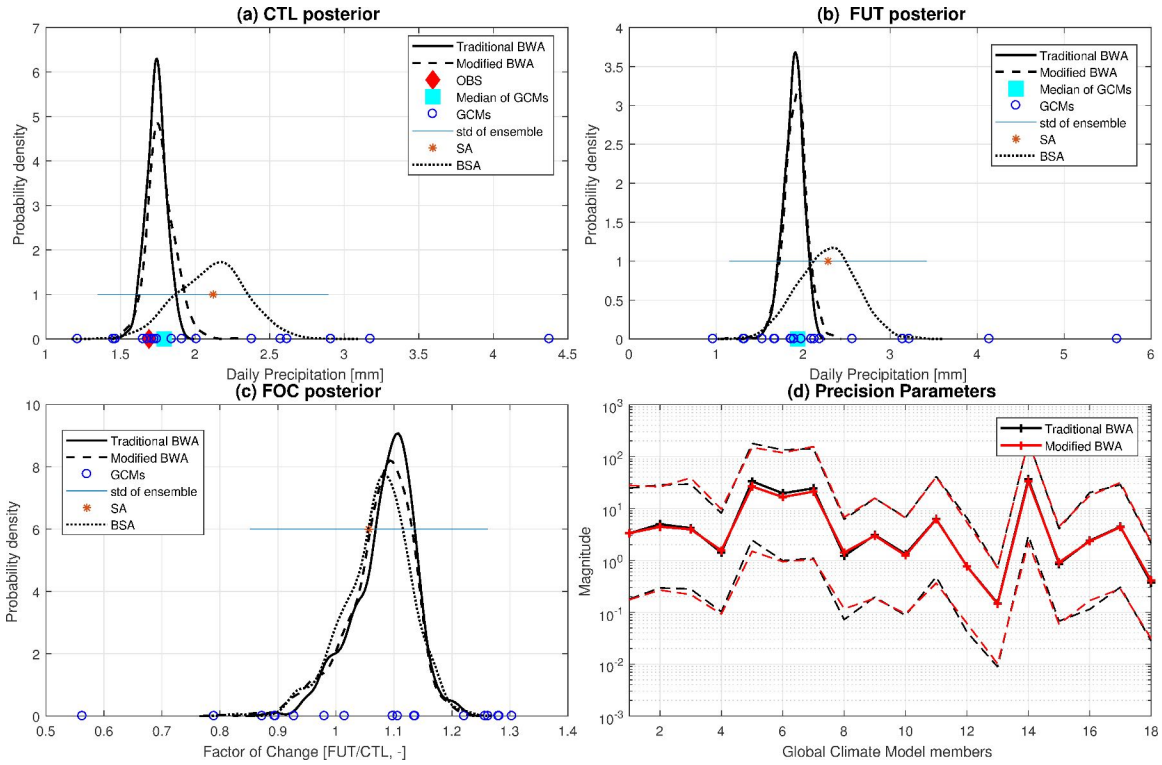

Figure 2.4: Posterior distributions, individual GCM results (circle symbols), and observations (red diamond symbol) for mean February precipitation for the location of Miami corresponding to **(a)** control period (1961-1990); **(b)** future period (2041-2070, CMIP5 RCP45 scenario); and **(c)** the PDF of factor of change [FUT/CTL]. Subplot **(d)** shows the median (solid lines) and the 5th-95th percentiles (lower and upper dashed lines) for GCM 'precision parameters' of *Tebaldi et al.* (2005) (λ_i) based on 1000 MCMC samples. 'Traditional BWA' indicates Bayesian Weighted Averaging method in *Tebaldi et al.* (2005); 'Modified BWA' indicates the method presented in this chapter; 'SA' is simpe averaging of ensemble of model projections; and 'BSA' indicates simple averaging, where constant, pre-determined model weights are used in Bayesian context. 'OBS' refers to the observation

unreasonable (Figure 2.6). This is because the posterior distribution for the future period is very close to zero due to the effect of observation. As a result, a significant part of the distribution extends into the negative part of the axis. The percentage of negative values in the MCMC samples from the joint distribution depends on the location of the mean parameters (i.e., $\tilde{\mu}$ and $\tilde{\nu}$) as well as model precision parameters, which account for the variance of the posterior distribution. To address the issue of physically unrealistic values for precipitation, we truncated the posterior distribution at zero and distributed the corresponding mass from the negative range uniformly to the positive side of the axis. Even though both methods have negative values in the posterior distribution due to low precipitation depth, the distribution estimated with modified method is not changed significantly after the truncation (Figure 2.6b, 0.2% of the samples are negative). Since the modified BWA only relies on the model convergence, this leads to a physically realistic range of precipitation values in the posterior distribution for dry months. The substantial uncertainty for the traditional BWA for the month of June (Figure 2.5c, 2.6c) is due to the product FOC is highly sensitivity when CTL precipitation is small: deviation of FUT estimates from the near-zero values for the CTL period leads to extremely large FOC, which are anyhow applied to very low magnitude of precipitation.

Figure 2.7 shows a comparison of the two methods for the location of Flint. Although Flint has an entirely different climate characteristic from the other two locations, the comparison exhibits features that are consistent with previous results. The posterior distributions of FOC for both BWA methods for this location agree with each other since GCMs projections are fairly consistent with observations, and the climate is relatively wet. As expected, the largest differences can be observed for the periods when the observation is outside the GCM ensemble.

It is important to note that in contrast to precipitation, GCM temperature estimates are more consistent with observation (Figure 2.2a) across the models and,

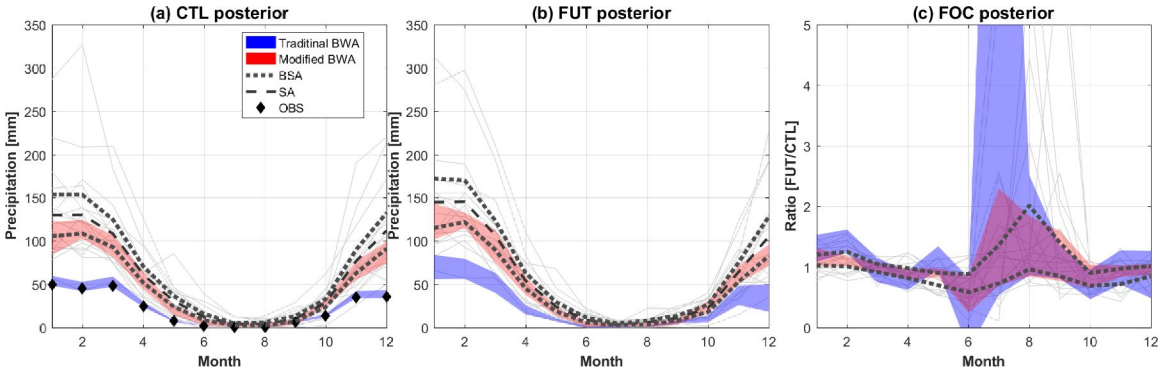

Figure 2.5: The 5th-95th percentile uncertainty bounds (blue and red) of the posterior distributions of mean precipitation for the location of Fresno corresponding to **(a)** control period (1961-1990); **(b)** future period (2041-2070, RCP45 scenario); and **(c)** the PDF of FOC [FUT/CTL]. FOC discontinuity of some GCMs is because the model output at CTL is averaged zero, and therefore the ratio of FUT and CTL is not defined

as has been pointed out in earlier studies, the primary source of uncertainty is the emission scenario (*Fatichi et al.*, 2016; *Hawkins and Sutton*, 2009; *Kim et al.*, 2016). This also applies to the case study locations (Figure 2.8): as seen, observations do not affect the estimation of FOCs significantly, and therefore revised version of BWA does not lead to any significant changes.

All simulation results for all of the considered scenarios for the three study locations (Miami, Fresno, and Flint), two emission scenarios (RCP45 and RCP85), and two time windows (mid-century, 2041-2070, and end-century, 2071-2100) can be found in Appendix A.

2.3.3 BWA versus BSA

Ensemble Simple Averaging (SA) and Simple Averaging in Bayesian (BSA) context are introduced to investigate whether the weighting scheme can appreciably affect the estimation of FOCs. Figure 2.1 illustrates that the mode of BSA posterior distribution for CTL and FUT periods is identical to the magnitudes obtained with the simple averaging method. This demonstrates that the definition of BSA's precision parameter in this chapter indeed leads to equal weights in the Bayesian context. Al-

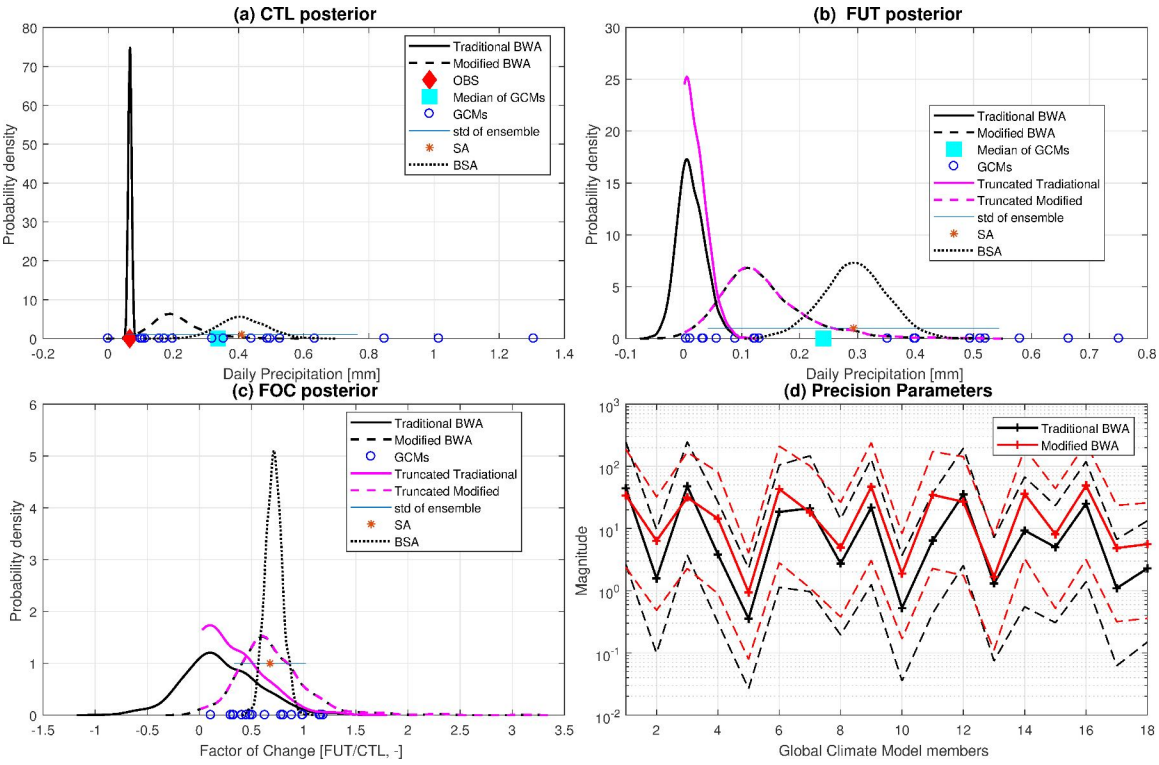

Figure 2.6: Posterior distributions (solid and dashed lines), individual GCM results (circle symbols), and observations (red diamond symbol) for mean June precipitation for the location of Fresno corresponding to **(a)** control period (1961-1990); **(b)** future period (2041-2070, CMIP5 RCP45 scenario); and **(c)** the PDF of factor of change [FUT/CTL]. Subplot **(d)** shows the median (solid lines) and the 5th-95th percentiles (lower and upper dashed lines) for GCM 'precision parameters' of *Tebaldi et al.* (2005) (λ_i) based on 1000 MCMC samples. 'Traditional BWA' indicates Bayesian Weighted Averaging method in *Tebaldi et al.* (2005); 'Modified BWA' indicates the method presented in this chapter; 'SA' is simple averaging of ensemble of model projections; and 'BSA' indicates simple averaging, where constant, pre-determined model weights are used in Bayesian context. 'OBS' refers to the observation. In subplots **(b),(c),** the PDF is truncated at zero and the mass is redistributed to positive values to maintain the area under the curve equal to 1.

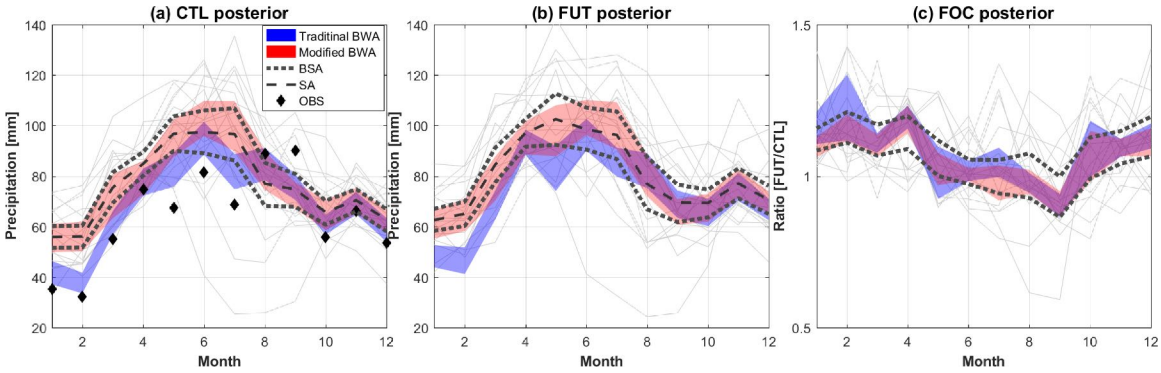

Figure 2.7: The 5th-95th percentile uncertainty bounds (blue and red) of the posterior distributions of mean precipitation for the location of Flint corresponding to **(a)** control period (1961-1990); **(b)** future period (2041-2070, RCP45 scenario); and **(c)** the PDF of FOC [FUT/CTL]. The light grey solid lines correspond to individual GCM monthly precipitation seasonalities. The two dotted dark grey lines are the 5th-95th percentile uncertainty bounds for the BSA method. 'OBS' refers to the observed data.

though *Fatichi et al.* (2013) previously concluded that in their case study weighting scheme does not significantly affect the determination of FOCs, we show that there are cases where a considerable difference exists among the posterior FOC distributions estimated with SA/BSA and the BWA methods. However, the results of the revised BWA approach are very close to that of BSA in terms of the first moment properties. In Figure 2.1c, the revised BWA has a smaller uncertainty and a small shift of the mean for both CTL and FUT as compared to BSA. This is because BSA is more sensitive to outliers since it considers all models being 'equal', while the revised BWA is more strongly constrained by the convergence of model outputs. Figures 2.3, 2.5, and 2.7 show quite similar comparison between revised BWA and BSA in terms of FOCs results for the three study locations. Even for the dry month (i.e., the June in Figure 2.6), no negative samples are generated from the MCMC process in BSA. This suggests that BSA may be a better alternate method for estimating FOCs when observed precipitation is pretty low.

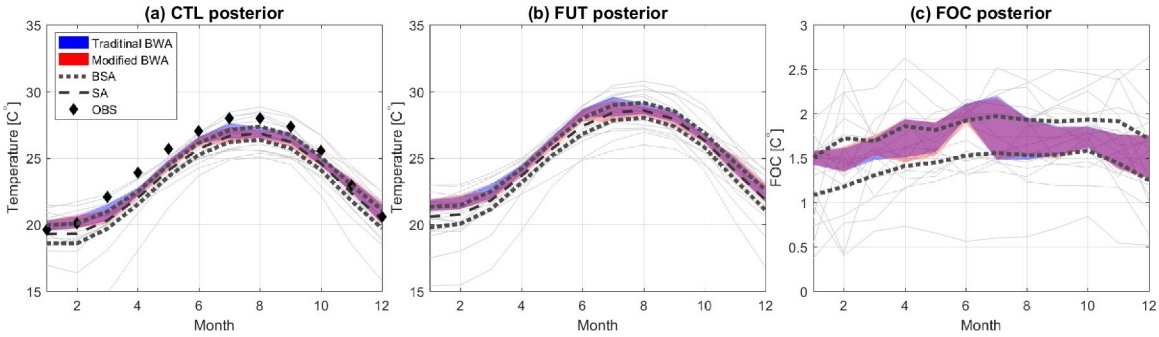

Figure 2.8: The 5th-95th percentile uncertainty bounds (blue and red) of the posterior distributions of mean temperature for the location of Miami corresponding to **(a)** control period (1961-1990); **(b)** future period (2041-2070, RCP45 scenario); and **(c)** the PDF of FOC [FUT/CTL]. The light grey solid lines correspond to individual GCM monthly precipitation seasonalities. The two dotted dark grey lines are the 5th-95th percentile uncertainty bounds for the BSA method. 'OBS' refers to the observed data.

2.4 Conclusions and discussion

A Bayesian averaging method that combines multiple model outputs based on their performance in terms of 'bias' and 'convergence' criteria is modified to exclude the former from the estimation of FOC from GCM projections, within the framework of a stochastic downscaling procedure. The modified approach is more robust in theory, as it is problematic to use the observation twice in the downscaling framework. Specifically, in the original method of *Tebaldi et al.* (2005), parameter λ_0 represents 'natural variability' for a climate variable (such as monthly precipitation) of the location of interest. This parameter, through the bias criterion, introduces observation-based weights in the estimation of the posterior distributions, which is typically much larger than performance-based weights introduced by individual models. An analysis of station-level data over the U.S. indicates, models rarely agree with observations strictly, and this is particularly true for precipitation (Figure 2.2). The implication is that an observational data point thus provides more information than any model output, with the consequence that the information content of the model ensemble has little weight. In the subsequent step of a typical downscaling procedure

(e.g., *Fatichi et al.* (2011)), the posterior distribution of FOC is applied to observation to derive future climate condition and the observation is assumed to represent CTL with 100% confidence in this stage. This procedure results in a theoretical inconsistency: observation information is doubly used with different confidence levels. It has a relatively large weight to account for the bias when combining outputs of the climate model ensemble to estimate FOC posterior distribution. However, it is used with a weight of '1' (e.g., certainty) when FOC is applied to derive the characteristics of future climate. The proposed modification eliminates this theoretical inconsistency by removing observation information in the first stage of the BWA process. Consequently, separating the estimation of FOCs from observation makes the downscaling framework less prone to artifacts.

Potential artifacts have been pinpointed in this study. The inclusion of observational data may introduce unpredictable impacts for precipitation downscaling. For a typical situation when precipitation observation deviates significantly from the model ensemble cluster, one or few models that happen to be closest to the observation receive the dominant weights in the estimation of the posterior distributions. When estimating distribution for future conditions, it may be dangerous to translate the model performance for the past conditions (bias criterion) into the model performance of the future conditions because of the weak relation between current/historic climate evaluation and future projection (*Knutti*, 2010). Further, the distribution of factors of change inferred using the original BWA approach may reflect a conflict of the two criteria defining individual model weight: the proximity of its output to historical observations (i.e., the bias) and the level of agreement with the other models (i.e., the convergence). As our results indicate, the two criteria can result in the bi-modality of the posterior distribution. The modified BWA leads to conceptually consistent posteriors that center around the model ensemble median. Another issue stemming from the inclusion of the bias criterion in the original BWA approach is the

presence of dry periods in site seasonality: physically inconsistent, negative values can be generated in the precipitation posterior distributions. While the revised BWA may also contain a negative region, its fraction is much smaller since the estimation relies only on GCM outputs, which are physically consistent.

To investigate the effect of the weighting scheme after removing observations from the original Bayesian framework, we compared the new-weighting scheme with the Bayesian Simple Averaging (BSA) approach that uses equal, pre-determined (before MCMC simulation) model weights. The median of the BSA posterior distributions is identical to the ensemble arithmetic average (referred to as SA here), but the spread of the BSA distribution is controlled by the precision parameter (see Eq (2.7)). Although there may be other formulations for the precision parameter, the solution of linking the spread of the posterior distribution to an estimate of the variance of model projections is a reasonable approximation. The difference between the results of the application of the original BWA and BSA/SA methods is non-negligible, while the modified BWA approach shows results comparable to BSA/SA. This comparison implies that the difference in weighting schemes is not crucial for determining FOCs when the bias criterion is not included. Although the validity of the convergence criterion as a metric of model credibility has been questioned (*Knutti*, 2010), removing the bias criterion appears to be reasonable for estimating FOCs and leads to a robust weighting method.

At last, we need to note that the weighting scheme is not critical for product FOC in very dry months since the absolute bias is negligible. In those cases, BSA shows superiority as compared to the modified BWA method, as no negative value is generated on the CTL, FUT, and FOC posterior distributions. Therefore, BSA may be an alternative method for estimating FOCs for dry climate conditions.

Additional results were developed for mean monthly temperature. A comparison between the original BWA method and the one proposed here shows that they are

very similar. The reason is that the accuracy of temperature modeling is higher than that of precipitation: temperature is a state variable of models, while precipitation is a flux much more challenging to predict. This is reflected in both better agreement of model outputs with observations (i.e., smaller bias, Figure 2.2) and lower coefficients of variation representing the variability of GCM ensemble (i.e., higher convergence).

In summary, we show that the criteria used to weight members of a multi-model ensemble can be critical in identifying the most probable climate change signal and the associated uncertainty and that observations must be used with care when one defines weighting criteria. A revised Bayesian Weighted Averaging method was proposed that retains key features of the original method. The revised approach accounts for the convergence of multiple models in simulating climate change signal and assigning uncertainties but avoids an explicit use of observations in estimating factors of change.

CHAPTER III

Change of Flooding Seasonality Caused by Climate Change

3.1 Introduction

Runoff is an essential source of fresh water, whose major phase is closely related to the life of humanity (*Hall et al.*, 2014). For example, the major phase of runoff can result in floods, which is among the most impactful natural hazards of all weather-related events in terms of fatalities and material costs (*Doocy et al.*, 2013). But it also replenishes reservoir and rivers, carries tremendous energy, and is an essential source of irrigation for agriculture in arid areas. Therefore, understanding the pattern of runoff in space and time is crucial for flooding control, water supplies, crop yield, water quality control, and hydropower generation. Its characteristics, such as magnitude, frequency, and seasonality, can be affected by human-induced climate change since warmer temperature intensifies the water cycle. The conclusions for the direction and magnitude of peak annual streamflow changes remain controversial between observation-based analysis and climate model projections (*Greve et al.*, 2018; *Gudmundsson et al.*, 2019; *Hirsch and Ryberg*, 2012; *Lins and Slack*, 2005; *Mallakpour and Villarini*, 2015; *Milly et al.*, 2005; *Yang et al.*, 2017). But there is high confidence that the frequency of extreme floods (associate with maximum runoff within a year)

has increased, and this trend will continue in the future (*Hirabayashi et al.*, 2013; *Hirsch and Archfield*, 2015; *Milly et al.*, 2002; *Slater and Villarini*, 2016). In terms of seasonality, numerous studies have been done to understand the effect of climate change on maximum annual streamflow occurrence date, so far they only have focused on historical period trends at station level (*Blöschl et al.*, 2017; *Clow*, 2010; *Cunderlik and Ouarda*, 2009; *Villarini*, 2016). In this study, we investigate the change of seasonality of peak runoff in the future at the US continental scale.

Strong flood seasonality has been reported over the US continent (see Figure 3.1), attributed to different flooding generation mechanisms (*Berghuijs et al.*, 2016; *Villarini*, 2016). Specifically, precipitation and soil water conditions are the factors explaining the occurrence of the highest flows over central US (*Slater and Villarini*, 2016) and west coastal areas (*Berghuijs et al.*, 2016; *Ye et al.*, 2017). In western mountain areas (*Li et al.*, 2017; *Yan et al.*, 2019) and northeastern US (*Hodgkins et al.*, 2003), snowmelt and snowfall are the dominant drivers of runoff. The higher temperature in the future can affect precipitation, soil moisture, snowmelt, which will result in a change of runoff seasonality in different regions.

In this study, we project changes in runoff seasonality for the end of the 21st century at the US continent scale based on runoff projections of 10 General Circular Models (GCM) from the fifth phase of the Coupled Model Intercomparison Project (CMIP5). GCMs estimate runoff (i.e., water excess in a grid cell), not streamflow (i.e., flow at a given point in the channel), and annual peak runoff is used to indicate the major phase of runoff in this work rather than annual peak flow (commonly used in other studies). Runoff routing to streamflow can introduce additional uncertainty since the coarse spatial resolution of GCM computational mesh cannot adequately represent basin lag. Further, a comparison between the *Livneh et al.* (2013) runoff dataset and streamflow measured at USGS gauges (see Figure 3.3) illustrates that the correlation between the average annual runoff and streamflow is strong in both

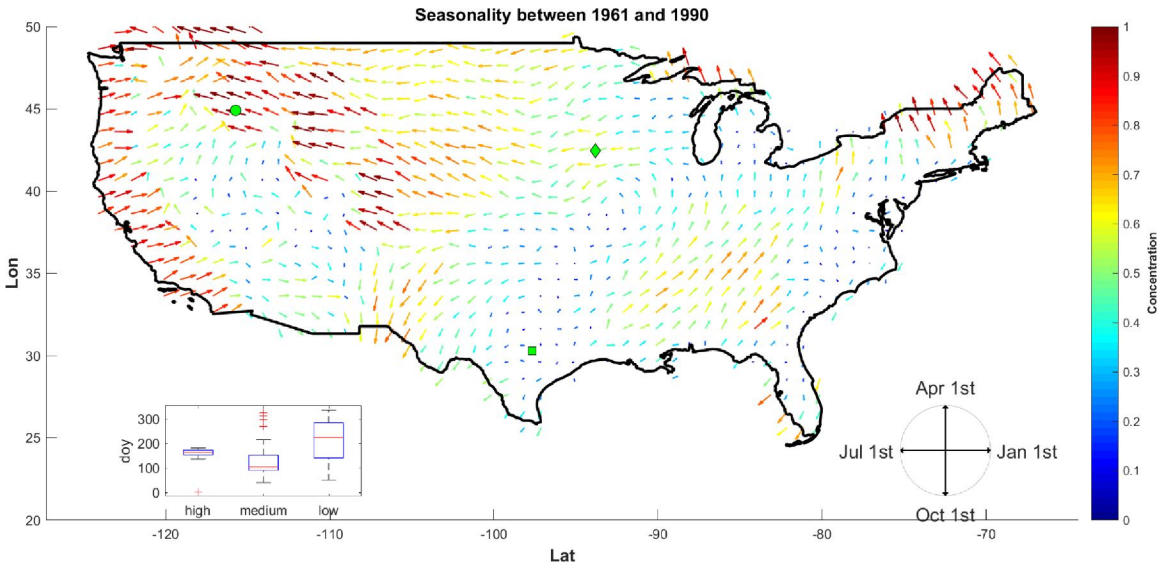

Figure 3.1: Peak annual runoff timing seasonality computed with daily runoff of *Livneh et al.* (2013) between 1961 and 1990. The direction of the arrow represents the averaged occurrence dates of 30 years, and the length and color represent the concentration metric (ψ) of *Blöschl et al.* (2017). The concentration metric (ψ) varies from 0 to 1, where $\psi = 1$ means the peak annual runoff occurs at the same date, where $\psi = 0$ means evenly distributed. 3 examples are selected to visualize the distribution of peak annual runoff dates in the inserted plot. The green circle, diamond, and square correspond to high ($\psi = 0.9$), medium ($\psi = 0.6$) and low ($\psi = 0.1$) concentration metric, respectively.

magnitude and timing, suggesting the lag effects of the watershed is not significant at this scale. Another advantage is that runoff projection allows us to study the spatial variability of the entire US continent excluding the effects of water management compared to point-level streamflow.

It has been established in the literature that projections based on a multi-model ensemble are preferred over single model outputs (*Knutti*, 2010; *Tebaldi and Knutti*, 2007). Biases of GCM projections of hydrological variables (e.g., temperature, precipitation) can be significant (*Knutti*, 2010; *Xu et al.*, 2018), and therefore they must be reduced before any robust conclusion climate change can be drawn. Consequently, we apply a Bayesian weighting averaging (BWA) method (*Smith et al.*, 2009) to produce multi-model ensemble inference, which relies on model performance in the past and convergence in the future to assign model weights.

3.2 Methods and Data

3.2.1 Multi-variate Bayesian Weighting Averaging

Bayesian weighted averaging (BWA) approach of *Smith et al.* (2009), *Tebaldi et al.* (2005) has grown in popularity as a sufficiently general tool to assess uncertainty in climate change from multiple model projections with minimum subjective assumptions. This approach is derived from the reliability ensemble average (REA) method introduced by *Giorgi and Mearns* (2002) to integrate outputs, such that the model weights are based on performance in the past period with historical observations and model output convergence in the future period. The first version of BWA was univariate such that each location was considered separately, creating solutions informed by the model performance at local scale (*Tebaldi et al.*, 2005). In the cases of substantial model - observation differences, this version could produce problematic posterior distributions (*Xu et al.*, 2018), also see Chapter II. To extend the approach utility, *Smith et al.* (2009) proposed a multivariate version of BWA that simultaneously considers a set of outputs in multiple regions. Model weights rely on its performance in all regions/locations considered, which ensures a more robust model skill evaluation given site-to-site variations of uncertainties. Specifically, the issues presented in Chapter II can be resolved with this multivariate version since the observation at a single location will not be overweighted. Another advantage of the multivariate Bayesian method is fewer parameters are required in calculating the posterior distribution than the univariate version. When a large number of locations are considered, using the multivariate Bayesian approach is much more computationally efficient. Readers are referred to *Smith et al.* (2009) for a detailed derivation, and only a brief description of the formulation is introduced here.

Smith et al. (2009) postulated that the jth climate model projections in the past and future at ith region are denoted as X_{ij} and Y_{ij}, e.g., $j = 1, \ldots, M$, and $i =$

$1, \ldots, R$, where M is the total number of models in an ensemble and R is the total number of regions considered. X_{i0} represents the associated historical observation from the same past period. It is assumed that observations and projections are random Gaussian variables that are distributed as:

$$X_{i0} \sim N\left[\mu_0 + \zeta_i, \lambda_{0i}^{-1}\right] \tag{3.1}$$

$$X_{ij} \sim N\left[\mu_0 + \zeta_i + \alpha_j, (\eta_{ij}\phi_i\lambda_j)^{-1}\right] \tag{3.2}$$

$$Y_{ij}|X_{ij} \sim N\left[\nu_0 + \zeta_i' + \alpha_j' + \beta_i(X_{ij} - \mu_0 - \zeta_i - \alpha_j), (\eta_{ij}\theta_i\lambda_j)^{-1}\right] \tag{3.3}$$

where λ_{0i} represents the variability of X_{i0} based on the observational data. Other parameters are assumed with the following prior distribution and all mutually independent:

$$\begin{gathered}
\mu_0, \nu_0, \zeta_i, \zeta_i', \beta_0, \beta_i \sim U(-\infty, \infty) \\
\theta_i, \phi_i, \psi_0, \theta_0, c, a_\lambda, b_\lambda \sim Ga\left[a, b\right] \\
\lambda_j|a, b_\lambda \sim Ga\left[a_\lambda, b_\lambda\right] \\
\eta_{ij}|c \sim Ga[c, c] \\
\alpha_j|\psi_0 \sim N\left[0, \psi_0^{-1}\right] \\
\alpha_j'|\alpha_j, \beta_0, \theta_0, \psi_0 \sim N\left[\beta_0\alpha_j, (\theta_0\psi_0)^{-1}\right]
\end{gathered} \tag{3.4}$$

Conventionally, $Ga\left[a, b\right]$ denotes gamma distribution with shape parameter a and rate parameter b. The parameters μ_0 and ν_0 are interpreted as the global means, ζ_i and ζ_i' are the differences from the global mean defined for a specific region 'i', and α_j and α_j' represent the global biases for a specific model 'j' for the past and future periods, respectively. In terms of the variance assumption in Eq (3.4), λ_j represents variation from the jth model, ϕ_i represents variation at ith region in the past, and ψ_i represents the variation at ith region in the future. The introduction of η_{ij} here is to guarantee climate models have different variation patterns at different regions. The uniform distribution is selected over $(-\infty, \infty)$, a, b, and c are set to 0.01 to

ensure all the priors are uninformative. Another three hyperparameters $\beta_0, \theta_0, \psi_0$ are used to define the common distribution of climate models. The analytical forms of the joint posterior distributions are unknown, but closed-forms of each marginal posterior distribution are derived in the appendix of *Smith et al.* (2009). In practice, Markov Chain Monte Carlo (MCMC) process is used to estimate the posterior distributions (*Smith et al.*, 2009).

The occurrence date of annual peak runoff is the variable of interest inferred from the GCM outputs. Daily GCM runoff outputs are used to derive the annual peak runoff timing, and Day of Year (DOY) is used to represent the occurrence date, with January 1st corresponding to 1 and December 31st to 365 (or 366 during a leap year). The original BWA cannot be applied directly to DOY due to its circular nature. For example, for a given grid location, the observed DOY of annual peak runoff is 5, and two climate models can produce outputs of annual peak runoff on days 10 and 365. Although both of such outputs correspond to the dates equally distant from the observation, the model with the value of 365 would be regarded as having a poor skill in BWA due to the large absolute difference of DOY values placed on a linear scale. To resolve this issue, we use the differences between the modeled and observed dates as the variable of interest in BWA to convert the circular variable to linear variable:

$$\begin{aligned} \tilde{X_{i0}} &= 0, \\ \tilde{X_{ij}} &= X_{ij} - X_{i0}, \\ \tilde{Y_{ij}} &= Y_{ij} = X_{i0}. \end{aligned} \tag{3.5}$$

$\tilde{X_{i0}}$ is taken as constant 0 for all the $i - s$ to evaluate model performance over the past period, since the more $\tilde{X_{ij}}$ deviates from 0 (e.g., $\tilde{X_{i0}}$), the less skillful for model X_{ij} to reproduce X_{i0}. A method of directional statistics (*Berens*, 2009) is used to calculate the differences in Eq. (3.5). Those new quantities are then used to replace X_{i0}, X_{ij}, Y_{ij} in Eq (3.3).

3.2.2 Robustness of GCMs

Although the Bayesian framework is used to increase the confidence of multi-model projections, the projection uncertainty can still be significant, particularly at a local scale (*Xu et al.*, 2018). To investigate the consistency of the GCMs used, a robustness metric (R) introduced in *Knutti and Sedláček* (2013) is used in this work:

$$R = 1 - \frac{A_1}{A_2}, \tag{3.6}$$

where A_1 is defined as the integral of the squared area between two cumulative density functions (CDFs), one constructed from combined individual model projections and the other one from the BWA projection for a future period. Variable A_2 is the integral of the squared area between two CDFs, both constructed from the BWA of model outputs for future and past periods. The A_1 and A_2 represent 'model noise' and 'model signal', respectively. Therefore, the robustness metric can be interpreted as a measure of model relative agreement on the change in future projections. In the original application of *Knutti and Sedláček* (2013), the Gaussian distribution is assumed for CDFs, and their parameters are estimated from the sample mean and variance. In this work, we use the von Mise distribution (*Abramowitz*, 1974) because of the circular nature of the downscaled DOY variables. The two parameters of the distribution, preferred direction (μ) and concentration (κ), are estimated from the samples with the circular statistics tool as in *Berens* (2009). Specifically, the parameters μ and $1/\kappa$ are analogous to the mean and variance in the normal distribution.

3.2.3 Runoff projections

Realizations from 10 GCMs developed in different institutions were downloaded from the CMIP5 database (`http://pcmdi9.llnl.gov/`) to reduce the dependence among models. The 10 GCMs selected in this study satisfy the criteria of availability

of daily runoff, precipitation, soil moisture, and snow amount outputs and completeness of spatial coverage over the contiguous US. Emission scenarios corresponding to Representative Concentration Pathway (RCP; *van Vuuren et al.*, 2011) 45 and 85 are used. RCPs are predictions of how concentrations of greenhouse gases in the atmosphere will change in the future as a result of human activities, such as RCP 45 represents moderate temperature increase, and RCP 85 assumes "business as usual" in the future. Three assessment periods were chosen: past (1961-1990), mid-century (2041-2070), and end-century (2071-2100).

3.2.4 Historical runoff dataset

Long-term estimates of daily runoff (surface water yield per unit area) provided by *Livneh et al.* (2013) is used in this study as true 'observations' within the Bayesian framework of multi-model downscaling to reduce bias. This daily runoff is obtained as output of the Variable Inltration Capacity (VIC) model (*Liang et al.*, 1994) forced with precipitation and temperature, at the spatial resolution of $1/16^\circ \times 1/16^\circ$.

We selected 5,217 daily streamflow gauges from United States Geological Survey (USGS; see Figure 3.2), which have at least 15 years data between 1961 and 1990, to verify the runoff dataset of *Livneh et al.* (2013). The majority of the watersheds are much smaller than the GCM resolution, therefore, it is not appropriate to correct the bias of GCMs runoff with streamflow observations. The accuracy of the average annual runoff magnitude and annual maximum runoff date is verified by the USGS streamflow data, as shown in Figure 3.3. The discharge time series of the USGS gauge data were converted to annual runoff by dividing by the respective contributing area (*Fekete et al.*, 2002). Runoff of all grid cells in *Livneh et al.* (2013) that are located inside a watershed corresponding to a given USGS gauged outlet were aggregated. The annual peak runoff dates from both datasets were subsequently identified and averaged for the selected period (e.g., 1961-1990) with the method of directional

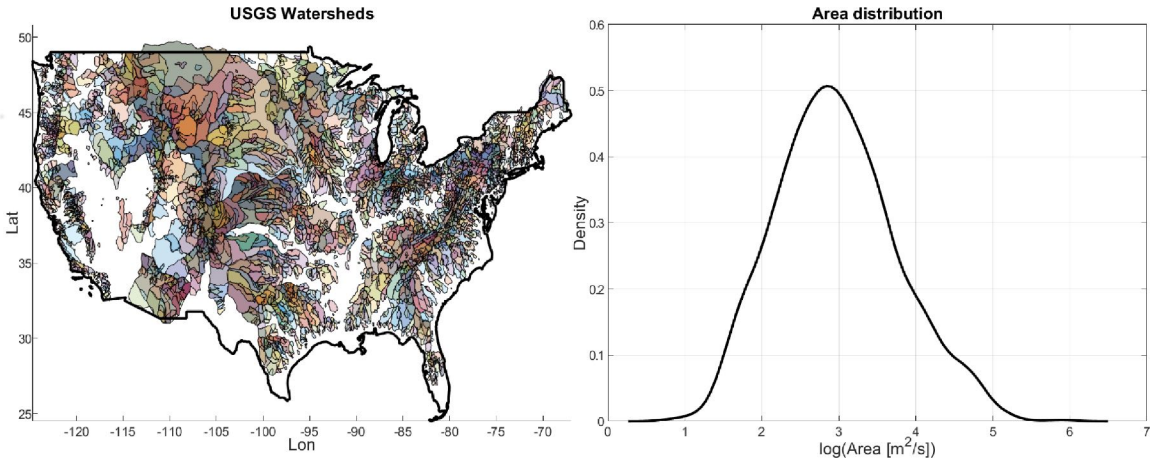

Figure 3.2: Watershed delineation of the selected 5,217 USGS daily streamflow gauges (left panel) and the distribution of the corresponding contributing area (right panel).

statistics (*Berens*, 2009).

Because GCM outputs and the runoff dataset have different meshes, they were converted to the same $1° \times 1°$ resolution for computational convenience. Specifically, we first re-mapped all GCM outputs to $1/16° \times 1/16°$ resolution with the nearest neighbor method. Then, both GCM and the runoff data layers were aggregated by averaging over grid cells falling inside the respective elements of the $1° \times 1°$ mesh.

3.3 Changes of peak annual runoff timing

BWA produces a posterior distribution of the multi-model mean of peak annual runoff dates for the past and future periods separately. Then, we can construct the posterior of the change based on the two posteriors. The change is defined as the distance of the peak runoff time between the future and historical periods. For example, Figure 3.4b shows the BWA posteriors of both periods at a single grid cell specified in subplot a, and subplot c shows the posterior of the change of the peak runoff seasonality.

Figure 3.5 illustrates the change of averaged peak annual runoff timing between the future and the control periods from multi-model BWA posterior at maximum

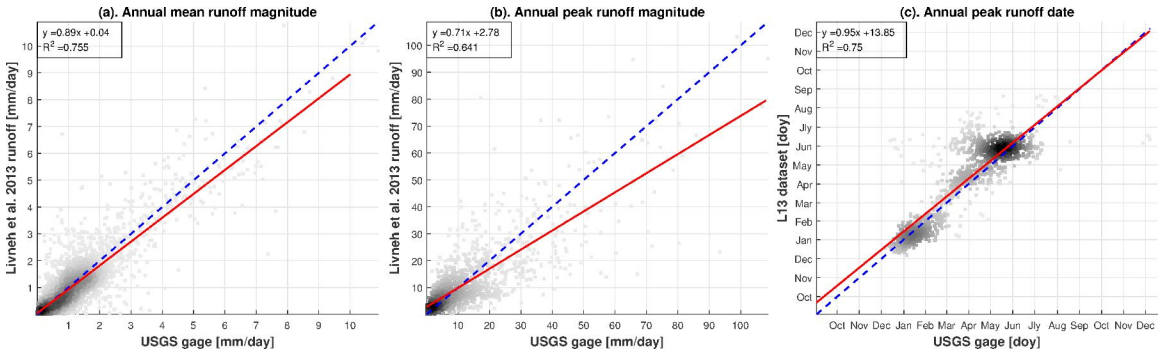

Figure 3.3: Validation of daily runoff data of *Livneh et al.* (2013) in terms of (a) mean annual runoff magnitude with all selected 5,217 USGS gauges, and (b) mean annual peak runoff magnitude , and (c) mean annual peak runoff date with 1,379 USGS gauges whose concentration metric of *Blöschl et al.* (2017) is larger than 0.6 over 1961 $\sim$ 1990. See Appendix B for how to the comparison is implemented. X-axis represents data based on observations at USGS gauges and Y-axis represents data product of *Livneh et al.* (2013). The blue dashed line is the 1:1 reference line and the red solid line is the linear regression line. The shading level depth is positively associated with the point density.

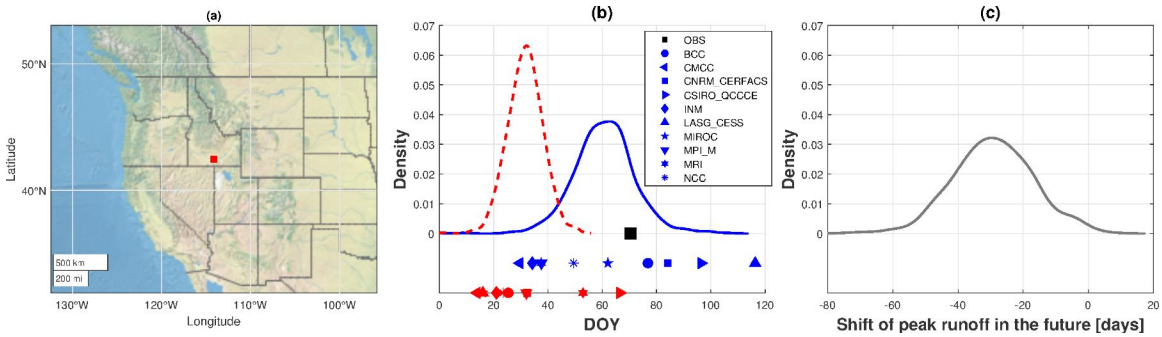

Figure 3.4: Multi-model BWA posterior of one grid cell locates in subplot (a) for the control period (solid blue line) and future period (end century, RCP85, red dashed line) in subplot (b). The blue markers represent the mean peak annual runoff dates from each GCM during the control period, and the red markers are for the future period. The black square is the mean date from *Livneh et al.* (2013) at this grid cell, which is used as "observation" in the Bayesian framework. Subplot (c) represents the posterior of the change of days between the future and past that constructed with the corresponding posteriors.

likelihood, and four cases due to the combination of two future periods and two emission scenarios are presented separately. Robustness metric of *Knutti and Sedláček* (2013) is used to calculate the robustness of the multi-model projections, and grids with high robustness (robustness metric $>$ 0.6) are stippled with green dots. The

fraction of areas showing changes with high robustness for these four time periods is (a). 9.3%, (b). 10.2%, (c). 10.7%, and (d). 17.2%, implying that the higher the increase in temperature, the more significant and consistent changes are projected by GCMs. In terms of spatial pattern, the changing trends are similar for all the cases for some regions: the snowmelt dominant regions, such as Rocky Mountain, and New England, are projected that annual peak runoff shifts to earlier months. The runoff peak is delayed for the Midwest, southern Florida, and part of the west coast, where the soil moisture is the key factor (*Ivancic and Shaw*, 2015). The change in the West Gulf exhibits high uncertainty due to the low concentration (e.g., peak annual runoff occurs uniformly over a year), where the monsoon typically causes the peak annual runoff. The changes have different directions in the upper Missouri basin between the end-of-century period for RCP85 and the other three cases, but model uncertainties can cause this.

3.4 Attributions for the change of peak annual runoff timing

To understand the attributions of this timing change pattern, we investigate outputs of daily precipitation, surface snow amount, and top layer soil moisture from CMIP5 multi-model ensemble. While extreme heavy precipitation (e.g., return period larger than 100 year) is always associated with peak annual runoff (*Smith et al.*, 2013), change of peak annual precipitation timing cannot explain the peak annual runoff timing change as shown in Figure 3.6. Specifically, there is no correlation between the change of maximum precipitation timing and the change of peak runoff timing. The maximum 3-day, 5-day, and 7-day accumulated precipitation are not associated with peak annual runoff based on our results (see Figure 3.11, 3.12, and 3.13). This is consistent with the studies based on stream gauge data showing that snowpack dynamic and antecedent soil wetness play a more critical role in generating peak runoff than precipitation only (*Ivancic and Shaw*, 2015), except for the urban

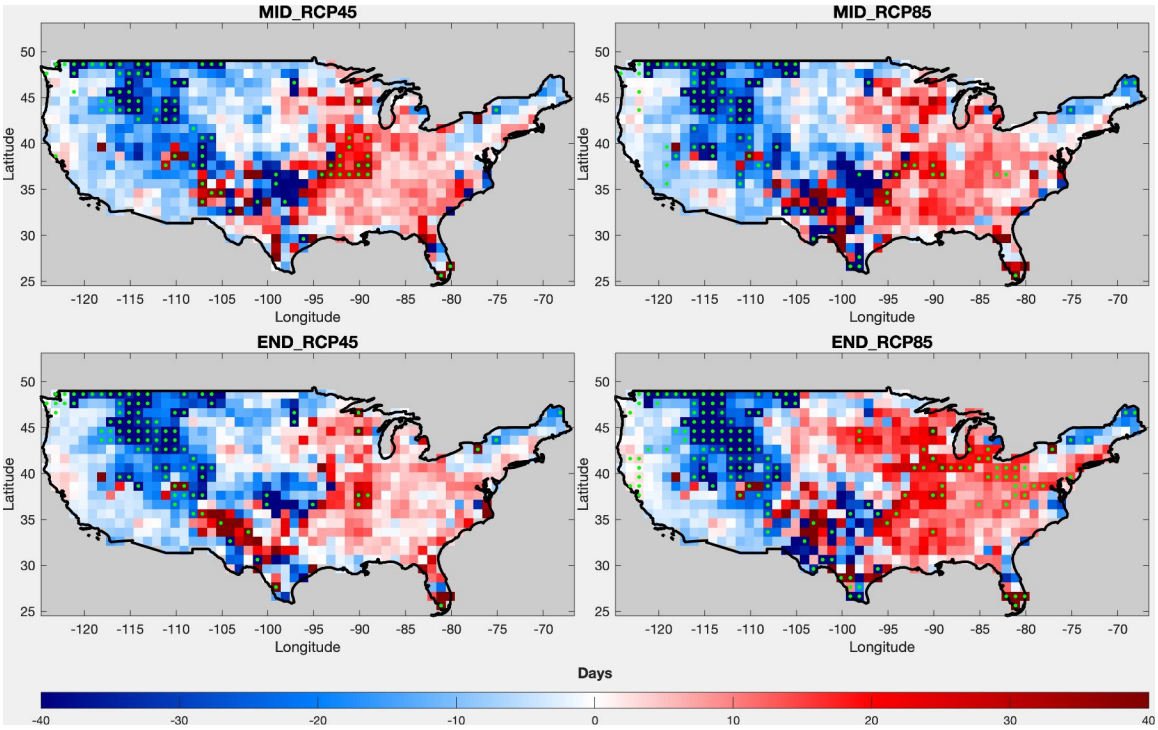

Figure 3.5: Change of mean peak annual runoff occurrence date between future period (FUT) and control period (CTL). The differences (FUT − CTL) are taken from the BWA posterior at maximum likelihood (Figure 3.4c). The grid cells with high robustness (metric of Knutti and Sedlek (2012) > 0.6) are stippled with green points.

areas where heavy rainfall is the primary driver (*Sharma et al.*, 2018).

We use maximum annual snow amount (i.e., the mass of surface snow) occurrence time to denote the onset of snow melting, and Figure 3.7 shows its change in terms of arithmetic multi-model mean. There is a strong correlation between the peak runoff timing change and the peak snow amount timing change for the grid cells with early shift annual peak runoff, and the corresponding robustness metric is larger than 0.6 (see Figure 3.10 blue squares). This good correlation supports the hypothesis that global warming induces earlier snowmelt and results in earlier peak runoff in the areas where snow-related processes dominate the runoff generation mechanism.

We selected projections of daily soil moisture of the top layer (at 10 cm) to investigate the relationship between soil wetness and peak runoff. Soil moisture can reach saturation many times in a given year, which makes it is different from pre-

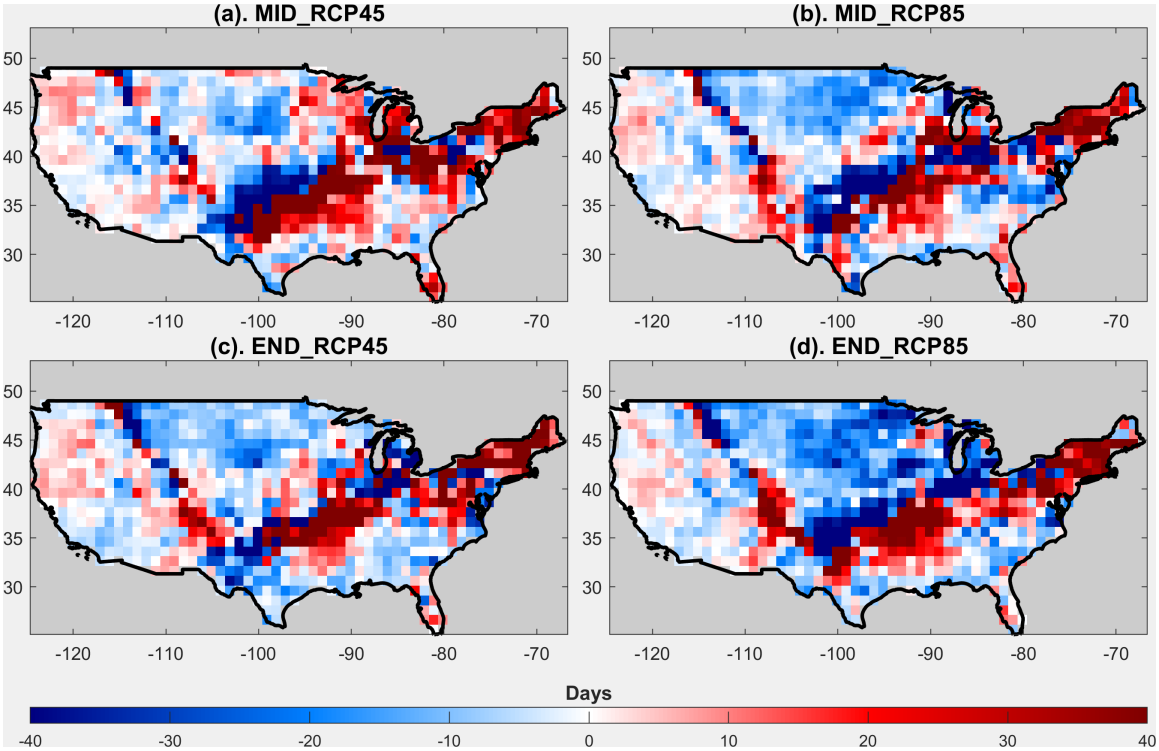

Figure 3.6: Change of annual peak precipitation time between future period and control period. The differences are calculated from the simple average mean of selected GCMs in terms of annual peak precipitation time (e.g., FUT CTL).

cipitation and snow amount that we can identify a single maximum value in a year. Thus, we identified all dates within the spring period (from February to May) when soil moisture is larger than 95% of saturated soil moisture out from the control period and future period separately from all the GCMs to construct cumulative density function (CDF). In this study, maximum soil moisture over all periods is identified as saturated soil moisture. Figure 3.8b shows the multi-model based CDFs of both periods for one of the grid cells that show delayed peak runoff corresponding to the end-century RCP 85 scenario, and the change is robust (e.g., grids stippled with green dots in Figure 3.5). Apparently, the CDF of future period shifts as a whole to a later time of the year compared to that of the control period, meaning that the centroid of the wet period during the spring is moving to a later time due to global warming. Since we constrain the dates in the CDF to the spring period, this delay reflects the

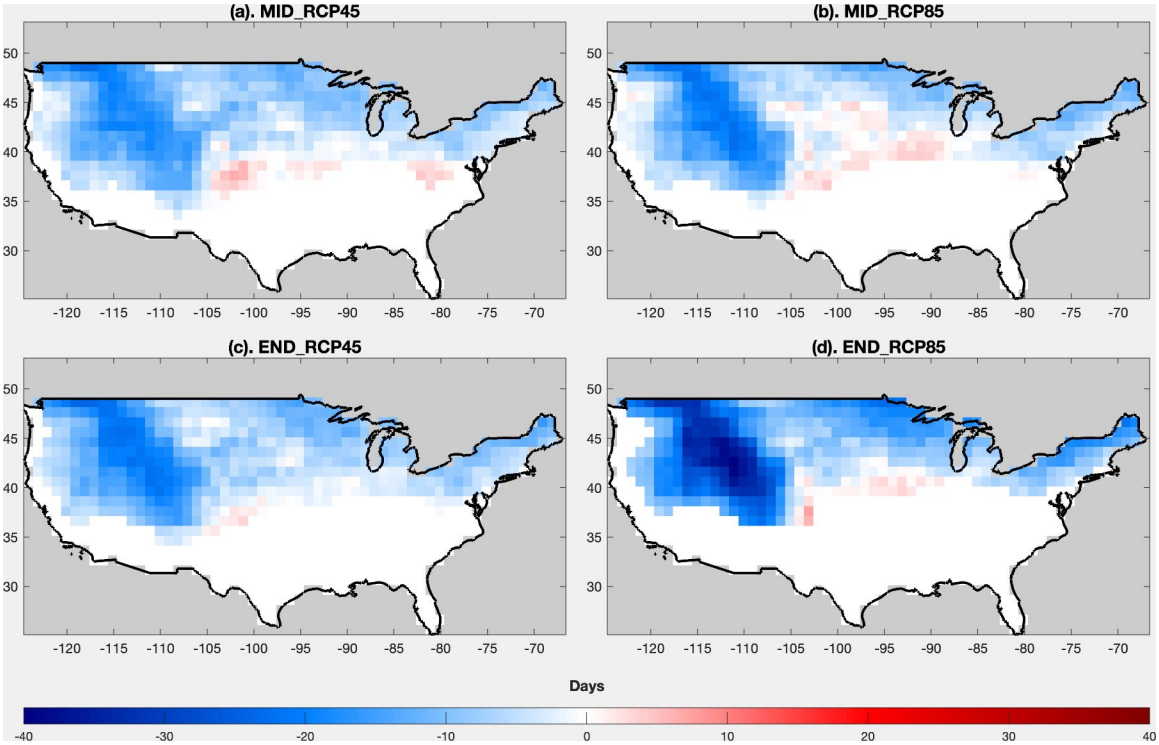

Figure 3.7: Change of annual maximum snow amount time between the future period and the control period. The differences are calculated from the simple average mean of annual maximum snow amount time (e.g., future - historical). The white pixels represent maximum snow amount is less than 100 kg/km^2 (equivalent to 10 mm in water depth) in the future period.

overall control of precipitation, evapotranspiration, and snow melts on the subsurface wetness due to the persistence of soil moisture (Ghannam et al. 2016). Peak annual runoff may occur on any day along with the CDF, but we calculate the difference between the two CDFs at the median value to assess the difference between the distribution centroids. Figure 3.9(d) illustrates this change over the whole US continent corresponding to the end-century RCP 85 scenario, which correlates with the shift of annual peak runoff timing clearly (see Figure 3.10). Additionally, as shown in Figure 3.10, the correlation is similar to the shift of CDF at 25% and 75% percentile, implying that the selection of the percentile on CDF does not affect the results.

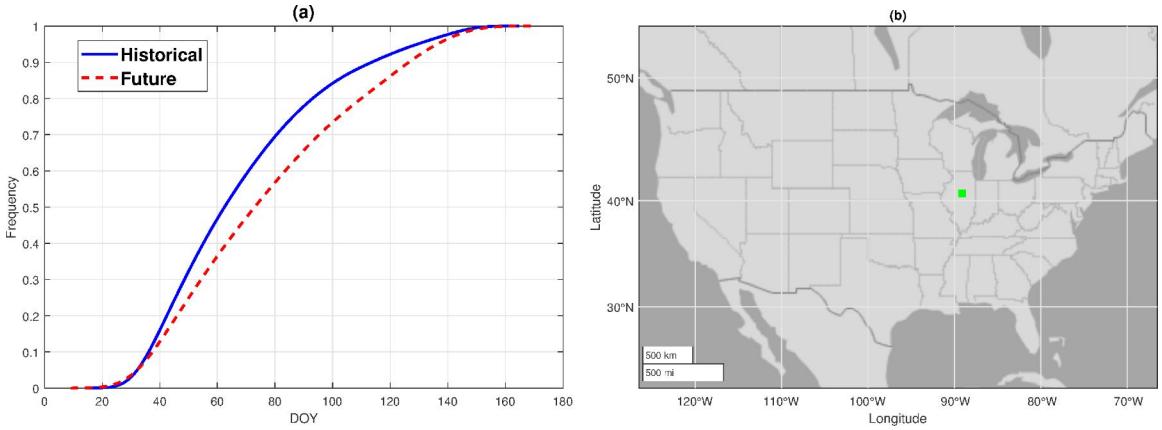

Figure 3.8: CDF of all the dates between February and May that soil moisture is larger than 95% of saturated soil moisture from all the selected GCMs during control period (blue solid line), and end century with RCP 85 (red dashed line) from the grid cell shown as green square in subplot (b).

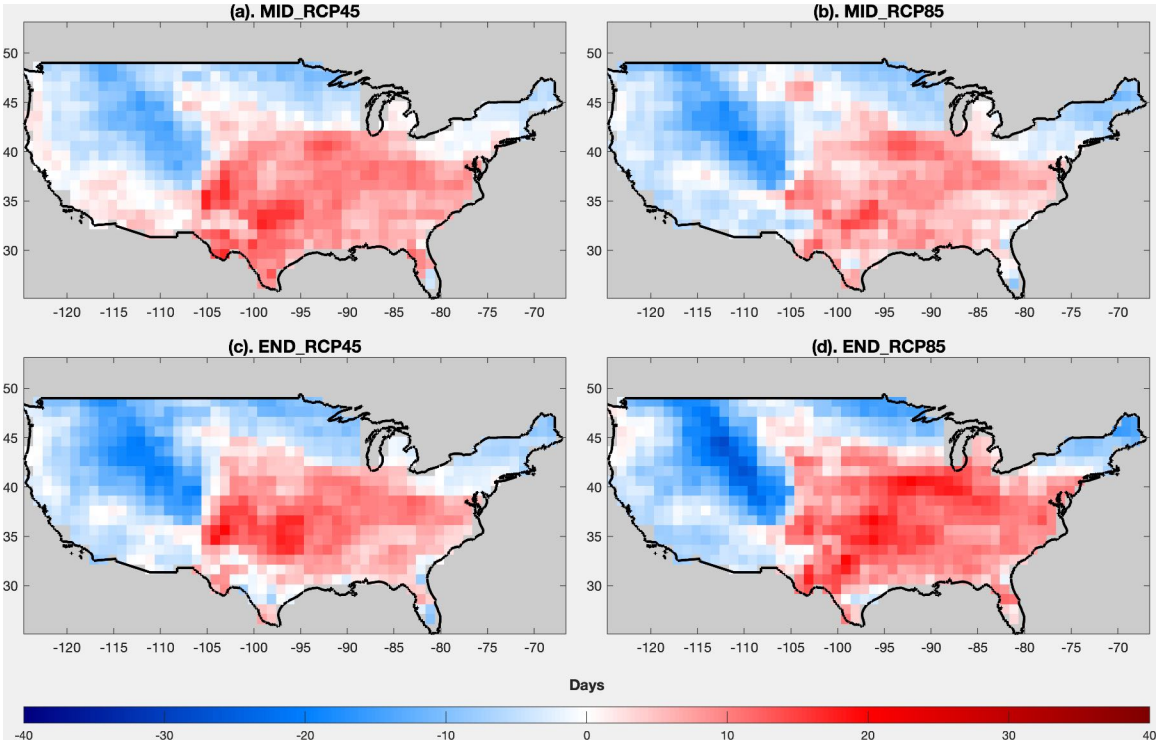

Figure 3.9: The shift of centroid of soil saturation period during Spring time (from February to May) between future period and control period. The soil saturation period includes the dates between February and May that soil moisture is larger than 95% of saturated soil moisture. We constructed CDFs for the soil saturation period with all the selected GCMs for the control and future period separately (see Figure 3.8 for the example of the CDFs at one grid cell). The values in the map of all the subplots are the differences between these two CDFs at 50% percentile.

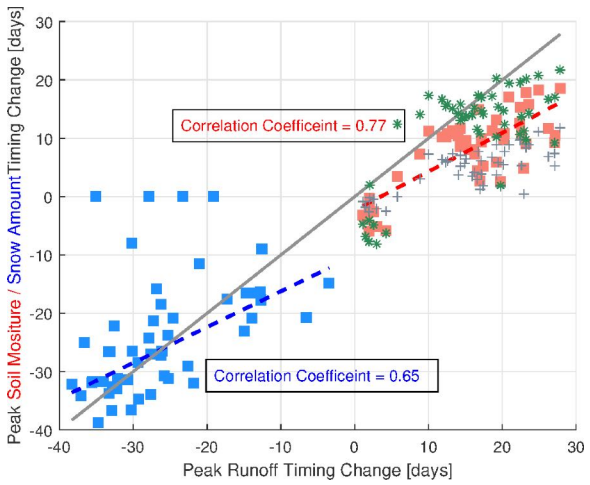

Figure 3.10: The regression between the peak annual runoff timing change and the peak snow amount timing change (blue squares), and the shift of median of CDF defined in Figure 3.8 for the end-century and RCP 85 scenario. Only pixels with early shift (delay shift) of peak annual runoff timing with high robustness metric are presented in the blue squares (red squares). The grey plus and star signs represent the regression of runoff timing change with the 25% (pluses) or 75% (stars) change of the distribution of spring days of soil saturation, respectively.

3.5 Conclusion

Our results show a clear pattern of peak annual runoff timing change caused by the global climate change that drives changes in the physical processes of land-surface hydrology. In summary, snowmelt will occur earlier in the future due to increased temperature, and this will cause an early shift of peak annual runoff in the regions where runoff is dominated by snowmelt. For the other regions with peak annual runoff timing projected to delay, we found the centroid of soil wetness period delays as well. Those changes can pose severe threats to the society and environment since we are already adapted to the original runoff seasonality (*Blöschl et al.*, 2017). Further, all the changes are projected to be more significant and more robust by the end of 21 century if we maintain the current greenhouse gas emission levels, as RCP 85 represents business as usual scenario (*van Vuuren et al.*, 2011). Hence, it is urgent for us to constraint the global warming under some level to mitigate the change of high runoff seasonality, for instance, the $1.5°C$ target set by Paris Agreement.

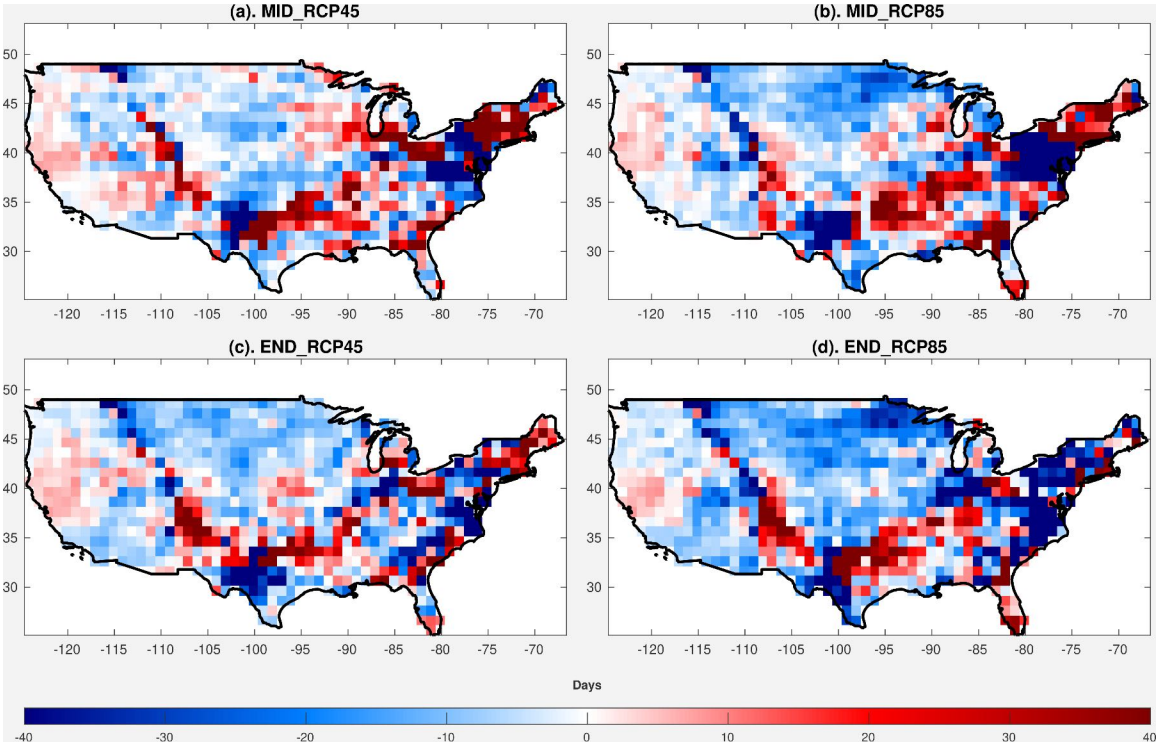

Figure 3.11: Change of annual maximum 3-day accumulated precipitation time between future period and control period. The differences are calculated from the simple average mean of selected GCMs in terms of annual peak precipitation time (e.g., FUT CTL).

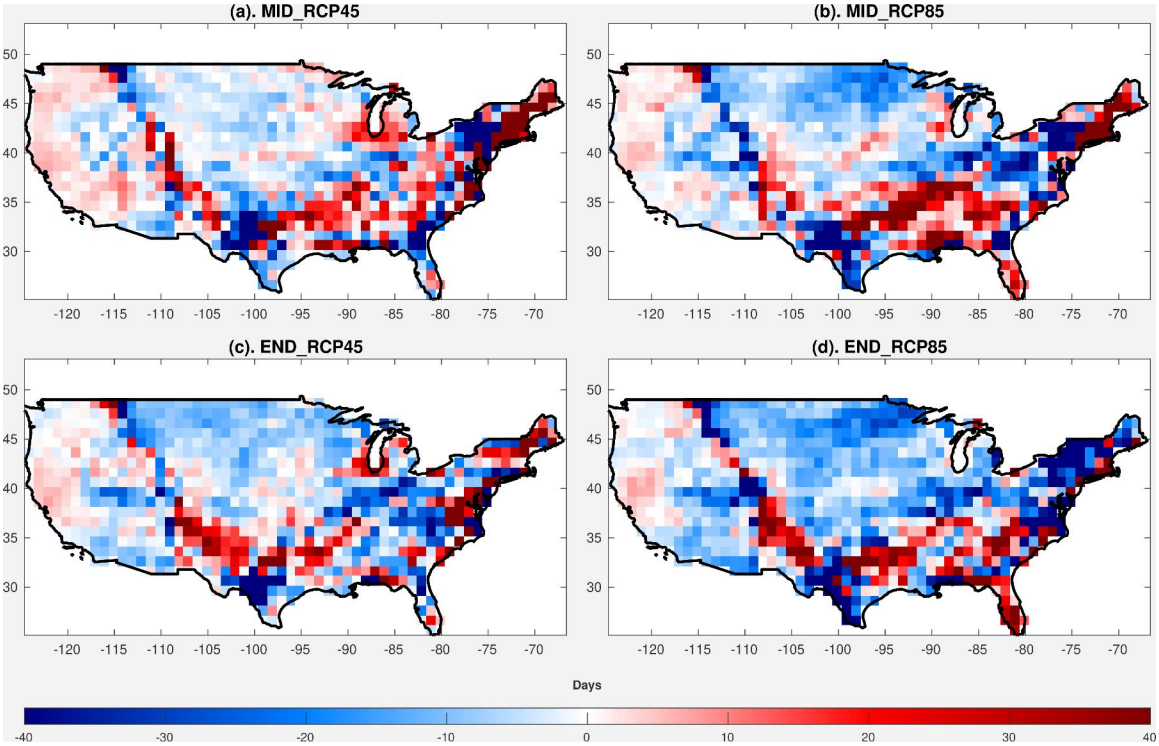

Figure 3.12: Change of annual maximum 5-day accumulated precipitation time between future period and control period. The differences are calculated from the simple average mean of selected GCMs in terms of annual peak precipitation time (e.g., FUT CTL).

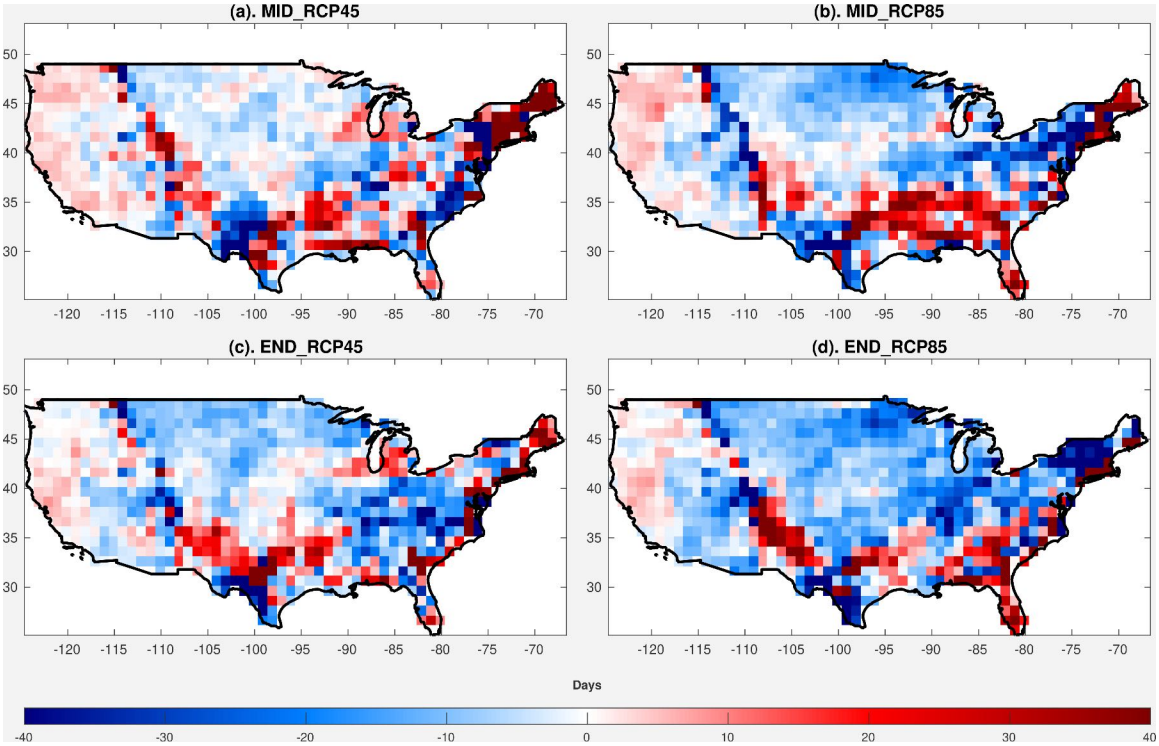

Figure 3.13: Change of annual maximum 7-day accumulated precipitation time between future period and control period. The differences are calculated from the simple average mean of selected GCMs in terms of annual peak precipitation time (e.g., FUT CTL).

CHAPTER IV

Estimation and Projections of Evapotranspiration Using Maximum Entropy Production Theory

4.1 Introduction

Evapotranspiration (ET) is an important component of surface water and energy cycle, which transfers the water from the land surface back to the atmosphere. ET accounts for 60% of the land precipitation (*Oki and Kanae*, 2006), and over half of the net radiation is used in this process (*Huang et al.*, 2017; *Trenberth et al.*, 2009) at the global scale. The temperature is projected to continue increasing until the end of the 21 century due to human activities, which has an impact on ET. Any changes in ET can affect precipitation, and the associated latent heat flux helps to control surface temperatures. Therefore, it is imperative to understand the effect of climate change on ET.

The reliability of ET estimation at any space-time scale remains challenging. Although eddy covariance (EC) measurement systems (*Gerken et al.*, 2018; *Restrepo-Coupe et al.*, 2013; *Baldocchi et al.*, 2001) provide observations of land surface fluxes at high temporal resolutions, they are representative of fairly small footprints ($\sim 1km^2$). Furthermore, EC measurement errors are generally as high as 20% (*Wilson et al.*, 2002), and pose certain gap-filling issues (e.g., during low turbulence mixing condi-

tions). Thus, models are required to simulate ET at any scale, using in-situ or remote sensing observations. Classical methods for estimating ET requires input data on water vapor gradient, wind speed, and surface roughness as well as model parameters such as stomatal resistance and atmospheric resistance. These inputs or parameters are often difficult to obtain on ground (*Brutsaert*, 1982), making it challenging to evaluate ET accurately in space and time.

Recently, a novel model based on the theory of maximum entropy production (MEP) (*Dewar*, 2005) was proposed by *Wang and Bras* (2009, 2011) for partitioning net radiation into latent (e.g., λE) or ET, sensible (H), and ground heat (G) fluxes. Unlike other methods that require ad hoc parameterizations, the MEP model only needs three inputs, i.e., net radiation, specific humidity, and temperature. The only parameter is soil thermal inertia of the surface where the radiative exchange takes place. This model has been validated with field observations for a diverse range of biomes (e.g., bare soil, sparse dry shrubs, grazed pasture, temperate forest, ocean, snow, and ice surface) with satisfactory performance (*Huang et al.*, 2017; *Nearing et al.*, 2012; *Shanafield et al.*, 2015; *Wang et al.*, 2014; *Yang et al.*, 2013). But the MEP model was not yet assessed for highly bio-diverse regions such as the Amazon to obtain ET estimates at regional scales. In this chapter, we first fill this gap by assessing whether the MEP model yields a robust skill across scales with inputs of different origins for the Amazon region.

4.1.1 ET estimation at Amazon rainforest

Amazon rainforest represents 60% of the world's remaining rainforests, playing a critical role in the global water, energy, and carbon cycles (*Davidson et al.*, 2012; *Kim et al.*, 2012b; *Malhi*, 2012). Amazon forest transfers a large volume of water from the land surface to the atmosphere through ET. Therefore, any change of ET over Amazonia may affect climate at the regional and global scales (*Maeda et al.*, 2017;

Spracklen et al., 2012). Due to the recent severe drought events in 2005 (*Marengo et al.*, 2008), 2010 (*Lewis et al.*, 2011), and the strong 2015-2016 El Niño episode (*Jiménez-Muñoz et al.*, 2016), it is imperative to investigate how the basin-wide ET might be affected, particularly if there is an impact from water stress.

Different methodologies have been applied to modeling ET over the Amazon rainforest. *Shi et al.* (2008); *Shuttleworth and Pereira* (1988); *Werth and Avissar* (2004) used the classical Penman-Monteith (P-M) formulation with canopy and aerodynamic resistances as the two key parameters. Canopy resistance behavior can be approximated through inverse estimation (*Jarvis et al.*, 1976), but effective parameterizations are cumbersome for tropical biomes with high diversity (*ter Steege et al.*, 2013). Other inputs, such as wind speed and surface roughness, are difficult to obtain. Bowen ratio energy balance method is another approach widely used to estimate ET (*Biudes et al.*, 2015; *de Abreu Sá et al.*, 1988; *Shi et al.*, 2008; *Sommer et al.*, 2002), however, the method is sensitive to the measurement accuracy of water vapor gradient (*Hu et al.*, 2014). Additionally, water budget analysis has also been applied to estimate ET at the Amazon basin scale (*Karam and Bras*, 2008; *Maeda et al.*, 2017). This approach computes ET as the residual of the basins water budget equation and requires precipitation, river discharge, and change in the basins water storage as input. Reliable historical discharge records are limited, preventing robust application of this method to most areas of the basin.

While previous studies carried out comprehensive analyses, environmental (e.g., net radiation, vapor pressure deficit) and biophysical (vegetation traits) controls on ET in the Amazon are yet to be fully understood (*Christoffersen et al.*, 2014; *Werth and Avissar*, 2004). For instance, net radiation is the main driving force of ET (*Fisher et al.*, 2009; *Hasler and Avissar*, 2007; *Hutyra et al.*, 2007). This is maintained even for the relatively dry southern part of the Amazon, where ET appears to be sustained by the soil water storage (*Jurez et al.*, 2007). However, General Circulation

Models (*Hasler and Avissar*, 2007) and land surface models (*Dickinson et al.*, 2006) report that annual ET is closely correlated with precipitation, implying that seasonal drought might have an appreciable control on the forest canopy conductance (*Costa et al.*, 2010). Hence, an interplay of the main ET limitations (energy vs. water) remains unclear for seasonally dry conditions and episodes of severe water stress. Furthermore, the role of leaf phenology in regulating forest water flux in the Amazon region during seasonal droughts is not well understood (*Christoffersen et al.*, 2014; *Manoli et al.*, 2018; *Wu et al.*, 2016). Overall, the estimation of water fluxes from the rainforest during drought periods requires an improved understanding of biophysical controls on canopy conductance, vegetation phenology, and root water uptake. The biodiversity of the Amazon region also implies that accurate parameterization of the primary ET agent – vegetation will remain challenging. Therefore, a parsimonious model of ET (i.e., a model with fewer inputs and parameters) is desirable, particularly at larger scales.

In this chapter, we apply the MEP method to a set of sites with detailed heat flux measurements and varying conditions of seasonal water stress. The MEP modeled ET is further compared with ET estimates from the widely used Moderate Resolution Imaging Spectroradiometer terrestrial ET (MOD16A2) product, at both local and basin scales.

4.1.2 Future projections of ET from GCMs

As discussed in Chapter II, model outputs from GCMs carry biases that need to be addressed by comparing to observations before any reliable statements can be drawn. However, the number of sites with measurements suitable for ET estimation is limited, and measurement periods are typically short compared to other variables (e.g., precipitation, temperature). Thus, the evaluation of GCM ET performance with observations is limited to sparse locations and short periods (*Yao et al.*, 2016). Other

studies use reanalysis, satellite-based products, Land Surface Model (LSM) results, or multi-dataset synthesis to assess ET estimates obtained from GCMs. (*Mueller et al.*, 2011, 2013; *Mueller and Seneviratne*, 2014).

In this chapter, we provide an assessment of ET estimates from GCM for the whole US continent with ET derived from the residuals using the water budget method. We use the precipitation and runoff provided by a long-term hydrological dataset of *Livneh et al.* (2013), which relies on Variable Infiltration Capacity (VIC) model driven by precipitation and temperature from numerous observed stations. Further, we show that the skill of GCM in simulating ET can be improved by running MEP model with net radiation, temperature, and specific humidity. Although MEP model requires fewer inputs than the classical model, it is still accurate since the information entropy guarantees the relevance and logical consistency of the information provided by the input variables (*Xu et al.*, 2019).

MEP model yields more robust projections of ET driven by GCM outputs, as the uncertainty of the variables that are hard to obtain (e.g., wind speed, roughness depth, resistance parameters) are excluded. Additionally, if the elevated CO2 will enhance the ET by increasing the area of transpiring leaves or suppress ET due to a decrease of stomatal conductance is still not clear (*Pan et al.*, 2015). The parsimony of MEP avoids the use of the stomatal resistance in estimating ET. Our results show that long-term annual ET performance at the US continent is improved for each GCM and multi-model ensemble by using the MEP method.

4.2 Data

4.2.1 Eddy flux sites

To verify MEP at the Amazon rainforest, we selected data from nine flux towers from the Large-Scale Biosphere-Atmosphere Experiment in Amazonia (LBA) project

(*Keller et al.*, 2004) to assess the modeled ET. A detailed description of site characteristics can be found in Table 1 of *Restrepo-Coupe et al.* (2013) and Table 4.1. The sites have the following abbreviations in this study: CAX - Cauxiana island (tropical rainforest), K34 Manaus (tropical rainforest), K67 Santarem K67 (tropical rainforest), K77 Santarem K77 (pasture-agriculture), K83 Santarem K83 (selectively logged tropical rainforest), BAN Bananal island (forest-savanna ecotone), RJA - Reserva Jaru (tropical wet and dry forest), FNS Fazenda Nossa Senhora (pasture), PDG - Pe de Gigante (savanna). Observations of latent and sensible heat fluxes are used in comparisons at their original (hourly) and aggregated (daily and monthly) temporal scales. Only days with valid measurements for all daytime hours (from 6:00 to 19:00) are used to compute heat fluxes at daily scale. The magnitude of nighttime ET is much lower than that of daytime flux, making it a negligible contribution, and therefore not considered in this analysis. Since the daily flux data may have gaps, only months with at least 20 days of available data are used for comparisons at the monthly scale. Using this criterion leads to sparse data for each individual site, therefore analyses at this scale involve multiple sites.

4.2.2 Drought index

Drought severity is represented by surrogate metrics and includes vapor pressure deficit (VPD) and cumulative water deficit (CWD). For a rainforest environment, VPD *indirectly* reflects the degree of soil control on the vegetation-mediated water flux that responds to the atmospheric demand (*Novick et al.*, 2016): the higher VPD, the higher the level of decoupling of the atmospheric state from the moisture source. VPD is computed for the height at which the eddy fluxes are measured (Table 4.1). CWD is an indicator of meteorologically-induced water stress used in several previous studies in the Amazon (*Aragão et al.*, 2007; *Lewis et al.*, 2011). More negative CWD corresponds to more severe water stress. Specifically, we use the precipitation data

Table 4.1: Eddy covariance sites used for validating MEP

ID	Site Name	Lat/Lon	Canopy Height [m]	Biome Type	LAI $[m^2/m^2]$	Measurement Period
CAX	Caxiuana	1.72S/51.53W	30-35	Tropical rainforest	5.14	01/01/99-07/30/03
K34	Manaus	2.61S/60.21W	30-35	Tropical rainforest	4.7	06/14/99-09/30/06
K67	Santarem K67	2.85S/54.97W	35-40	Tropical rainforest	6.0	01/02/02-31/12/14
K77	Santarem K77	2.42S/54.88W	0-0.6	Pasture-Agriculture (since Dec 2001)	2.52	01/01/00-12/30/05
K83	Santarem K83	3.01S/54.58W	35-40	Selectively logged tropical rainforest	4.9	06/29/00-03/12/04
BAN	Bananal island	9.82S/50.13W	18	trees understory Seasonally flooded,forest-savanna ecotone	3.5-4.5	10/24/04-12/08/06
RJA	Reserva Jaru	10.08S/61.93W	30	Tropical wet and dry forest	5.5	03/12/99-11/14/02
FNS	Fazenda Nossa Senhora	10.76S/62.36W	0.2-0.5	Pasture	1.4(2000) 2.8(2003)	02/04/99-11/04/02
PDJ	Pe de Gigante	21.62S/47.63W	1-3 (73%) ¡8(27%)	Savanna (cerradao)	6	01/01/04-12/31/06

from Tropical Rainfall Measuring Mission (TRMM), and estimate averaged evapotranspiration from the flux tower data for wet and dry seasons to construct CWD. The following algorithm is used to generate CWD time series for month 'n':

$$\begin{aligned} &CWD_n = CWD_{n-1} - ET_n + P_n, \\ &if \quad CWD_n > 0, \quad then \quad CWD_n = 0, \\ &CWD_0 = 0 \end{aligned} \tag{4.1}$$

Since the calculation of CWD should start from the wet period with CWD = 0 and the Amazon basin experienced intense drought in 1997-1998 (*Williamson et al.*, 2000), we compute monthly CWD from 1999 to 2017.

4.2.3 Remote sensing and reanalysis products

MOD16A2 product provides global terrestrial ET using a modified P-M method (*Mu et al.*, 2007, 2009, 2011) driven by remote sensing data from MODIS and the global reanalysis meteorological data from the Modern-Era Retrospective Analysis for Research and Applications (MERRA; *Rienecker et al.*, 2011):

$$ET = \frac{\Delta(R_n - G) + \rho c_p(e_s - e_a/r_a)}{\Delta + \gamma(1 + r_s/r_a)} \tag{4.2}$$

where ρ is the air density, c_p is the specific heat of air, e_s is the saturated vapor pressure, e_a is the actual vapor pressure, Δ is the slope of vapor pressure curve, γ is the psychrometric constant, r_a is the aerodynamic resistance, and r_s is the surface resistance. Collection 5 provides continuous ET at a temporal resolution of 8 days (available through Google Earth Engine `http://earthengine.google.com`) or monthly (available through Numerical Terradynamic Simulative Group, NTSG, `http://www.ntsg.umt.edu/`), both at 1 km resolution. MOD16A2 Collection 6 data product is also available but not used here due to lower spatial coverage and similar

performance as collection 5.

To estimate the basin scale ET over the Amazon, net radiation, and surface temperature from the SYN-1deg monthly Clouds and the Earth's Radiant Energy System (CERES) product (*Wielicki et al.*, 1996), as well as the MERRA product, are used to drive the MEP model. The finer 3-hourly SYN-1deg CERES product was previously used by *Huang et al.* (2017) to produce global ET. Since direct observations of ET do not exist, the MEP-simulated ET at the basin scale was assessed against MODIS ET product and water budget analysis. The original 1-km resolution MODIS product was converted to the 1-degree resolution by averaging all MODIS grid values inside the corresponding CERES grid cells.

4.2.4 Water budget method

Long-term mean annual ET can be obtained from the water balance equation:

$$ET = P - R + \frac{dS}{dt} \tag{4.3}$$

where P is the mean annual precipitation , R is the mean annual runoff, and $\frac{dS}{dt}$ represents the rate of change of soil water storage. For evaluating averaged ET at the amazon basin scale, mean precipitation from TRMM 3B42 daily product (*Huffman et al.*, 2007) aggregated to the basin scale and the mean annual runoff obtained by dividing streamflow at the most downstream gage station Obidos (`http://www.grdc.sr.unh.edu/html/Polygons/P3629000.html`) by the contributing area is used in Eq (4.3). The streamflow data are provided in the supplementary material of *Lopes et al.* (2016). While the watershed area at Obidos does not correspond to the entire Amazon basin (see its delineation in Figure 4.1), the analyses are carried out for this catchment. The data for the period of 2003 - 2013 were used here, assuming that the change of water storage is negligible (i.e., $\frac{dS}{dt} \approx 0$).

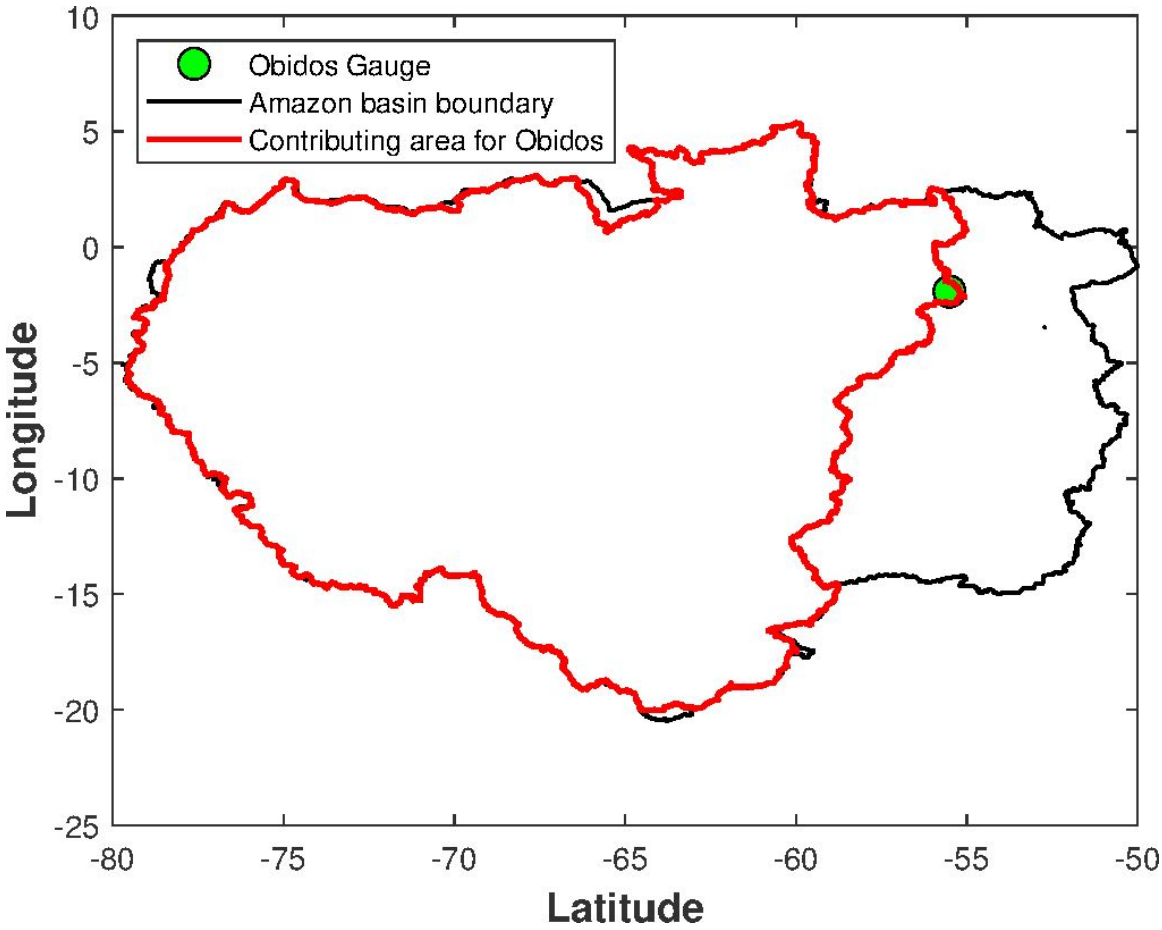

Figure 4.1: Boundary for the Amazon basin and the contributing area for streamflow gauge at Obidos

For the analysis of ET at the US continent, annual mean precipitation and runoff between 1961 and 1990 are taken from *Livneh et al.* (2013). The change of water storage in the soil is neglected here as 30 years is long enough. Due to the lack of observations of ET for the selected period, the performance of this ET estimation is validated indirectly by comparing the precipitation (see Appendix B) and runoff with WebMET precipitation and USGS streamflow (see Figure 3.3).

4.3 Maximum Entropy Production theory

Readers should refer to *Wang and Bras* (2009, 2011) for a detailed derivation of the MEP model, and *Hajji et al.* (2018) for the improvement of model performance

under water stress by introducing a stomatal openness related parameter. Only a brief description of the MEP model is shown in the following.

Net radiation (R_n) is partitioned into surface heat fluxes λE, H, and G as:

$$R_n = \lambda E + H + G \tag{4.4}$$

In the MEP theory, a dissipation function of energy fluxes D is expressed as:

$$D(\lambda E, H, G) \equiv \frac{2\lambda E^2}{I_e} + \frac{2H^2}{I_a} + \frac{2G^2}{I_s} \tag{4.5}$$

Where I_e, I_a, and I_s are the thermal inertia associated with the corresponding fluxes described in *Wang and Bras* (2009, 2011). There exists a unique solution for each flux term in Eq (4.5) that maximize the dissipation function of Eq (4.5) under the constraint of energy conservation of Eq (4.4) (*Wang and Bras*, 2010):

$$\begin{aligned} G &= \frac{B(\sigma)}{\sigma}\frac{I_s}{I_0}H|H|^{-\frac{1}{6}} \\ \lambda E &= B(\sigma)H \end{aligned} \tag{4.6}$$

where $B(\sigma)$ is a function of a dimensionless parameter σ characterizing water and thermal state of the evaporating surface (*Wang and Bras*, 2011):

$$\begin{aligned} B(\sigma) &= 6\left(\sqrt{1+\frac{11}{36}\sigma}-1\right) \\ \sigma &= \frac{\lambda^2}{c_p R_v}\frac{q_s}{T_s^2} \end{aligned} \tag{4.7}$$

In the above, λ is the vaporization heat of liquid water, R_v is the gas constant for water vapor and q_s and T_s represent the specific humidity and temperature at the transpiring surface, respectively. Combining Eqs. (4.4), (4.6), and (4.7), one can solve for λE (and H). Two versions of the MEP formulation are presented in *Wang*

and Bras (2011): soil evaporation and canopy transpiration.

4.3.1 Evaporation

Only evaporation (Ev_{MEP}) from the soil is considered when the land surface is bare soil. The thermal inertia of soil is related to the soil moisture:

$$I_s = I_{ds} + \sqrt{\theta} I_w \tag{4.8}$$

In the above, I_{ds} is the thermal inertia of dry soil, θ is the volumetric soil moisture, and I_w is the thermal inertial of water (e.g., 1.56×10^3 $[Jm^2K^1s^{120}]$). For the dominant types of soils on Earth, I_{ds} varies from 600 to 1000 $[Jm^2K^1s^{12}]$ (*Farouki et al.*, 1981; *Wang et al.*, 2010). A constant value $I_{ds} = 800$ $[Jm^2K^1s^{12}]$ is used in this study, and *Huang et al.* (2017) suggests such assumption will not introduce significant uncertainty for the ET estimation. Parameterization of I_a, I_e, and I_0 can be found in *Wang and Bras* (2009, 2011). Further, q_s and T_s should be measured at the soil surface to compute σ.

4.3.2 Transpiration

For the case of closed canopy where the energy balance is defined at the canopy top, ground heat flux G is negligible compared to λE and H. Thus, the energy conservation of Eq (4.4) is reduced to:

$$R_n = \lambda E + H \tag{4.9}$$

Under this constraint, the solution of Tr that maximizing Eq (4.5) in the MEP context is:

$$Tr = \frac{R_n}{1 + B^{-1}(\sigma)} \tag{4.10}$$

Recalling the equation of $B(\sigma)$ in Eq (4.7), Tr can be solved. In this version, q_s and T_s represent the specific humidity and temperature at the leaf surface, respectively. Those measurements at the leaf level are commonly unavailable. Since MEP model is insensitive to temperature (*Xu et al.*, 2019), T_s may be replaced with near-surface air temperature (T_a) without leading to appreciable uncertainty.

When unavailable, q_s my be parameterized in terms of specific humiidty within leaf stomatal cavities $q^{sat}(T_s, \psi_s)$ as:

$$q_s = \eta_s q^{sat}(T_s, \psi_s) \tag{4.11}$$

where $q^{sat}(T_s, \psi_s)$ is given in Eq (A2) of *Wang and Bras* (2011), ψ_s is leaf water potential, and $\eta_s (0 \leq \eta_s \leq 1)$ is a parameter related to the openness of leaf stomatal apertures. Specifically, when stomata close to prevent water loss, this results in a difference between q_s and q^{sat} within the stomatal cavity, expected to be particularly pronounced during water stress periods. *Hajji et al.* (2018) used soil moisture to evaluate η_s and showed an improved MEP performace under significant water stress:

$$\eta_s(\theta) = min\left[1, \frac{10(\theta - \theta_{min})}{3(\theta_{max} - \theta_{min})}\right] \tag{4.12}$$

where θ is the volumetric soil moisture, θ_{min} and θ_{max} represent the soil moisture of capacity and wilting point. In practice, θ_{min} and θ_{max} can be estimated by the 1st and 99th percentile of the distribution of long-term soil moisture data when the soil types are unknown.

However, the degree to which water stress controls transpiration in Amazonia is still unclear (*Christoffersen et al.*, 2014), making it challenging to determine η_s. When estimating ET in Amazon basin, above-canopy air specific humidity (q_a) was used as a surrogate of q_s in the site-level analysis since surface soil moisture is unlikely to be a suitable metric to characterize water stress of the Amazon where plants may take

up water from very deep soil (*Brum et al.*, 2019; *Ivanov et al.*, 2012; *Nepstad et al.*, 1994). In the MEP-based analysis that relies on CERES product, q_s is replaced with q^{sat}.

4.3.3 Vegetation fraction

ET of any GCM grid consists of evaporation from bare soil and transpiration from the vegetated surface, therefore, the vegetation fraction is required to separate ET. *Hajji et al.* (2018) proposed a relationship between Normalized difference vegetation index (NDVI) and vegetation fraction in the context of MEP theory:

$$f_{veg} = \frac{NDVI - NDVI_{min}}{NDVI_{max} - NDVI_{min}}. \tag{4.13}$$

However, NDVI is not available in GCM and any other satellite dataset during the control period. Leaf area index (LAI) is a dimensionless quantity that characterizes plant canopies. It is a measure for the total area of leaves per unit ground area and directly related to the amount of light that can be intercepted by plants. We use Leaf Area Index (LAI) as a surrogate of NDVI to approximate vegetation fraction since there exists a relationship between LAI and NDVI (*Saito et al.*, 2001):

$$f_{veg} = \frac{LAI - LAI_{min}}{LAI_{max} - LAI_{min}}. \tag{4.14}$$

Once f_{veg}, Ev_{MEP}, and Tr_{MEP} is calculated at each grid cell, the ET of MEP is:

$$\lambda E_{MEP} = (1 - f_{veg}) \times Ev_{MEP} + f_{veg} \times Tr_{MEP} \tag{4.15}$$

where the subscript "MEP" refers to the MEP method for estimating evapotranspiration, as given above in Eq. (4.10) for Tr and Eq. (4.6) for Ev.

4.4 Energy budget imbalance

4.4.1 Bowen Ratio Closure method

A typical magnitude of measurement errors of the EC method is 20% of the budget (*Gerken et al.*, 2018; *Wilson et al.*, 2002), leading to imbalance of the measured surface energy budget. The imbalance issue makes it difficult to evaluate model performance by directly comparing the model estimates with EC surface flux data. Specifically, in the context of the MEP approach, the energy conservation statement as in Eq (4.9) is automatically satisfied, while the sum of EC-measured latent and sensible heat flux is often lower than R_n as shown in Figure 4.2 and 4.3 for hourly and daily scale, respectively. The Bowen ratio closure method (*Mallick et al.*, 2016; *Twine et al.*, 2000) is used here to address this problem. Bowen Ratio (Bo) obtained directly from the measurements of λE_{OBS} and H_{OBS} is assumed to be valid, and the net energy imbalance is redistributed to the surface heat fluxes:

$$\begin{aligned} Bo &= \frac{H_{OBS}}{\lambda E_{OBS}} \\ \tilde{\lambda E} &= \frac{R_{nOBS}}{1+Bo} \end{aligned} \tag{4.16}$$

where OBS refers to observed fluxes, and $\tilde{\lambda E}$ is the corrected flux. The assumption of invariance of the Bowen ratio in Eq (4.16) implies a specific relationship between the errors of heat flux measurements, demonstrated here to be acceptable. Furthermore, one needs to keep in mind that Eq (4.9) is a reduced form of the energy budget equation for a vegetated surface (*Brutsaert*, 1982, p. 209). Diurnal dynamics of heat storage in the biomass and air and diffusive ground heat flux are not accounted for as they are believed to represent relatively smaller energy budget components during daylight hours for a rainforest environment (*de Abreu Sá et al.*, 1988). Ground heat flux can be safely assumed to be negligible for the majority of the sites, with the possible exception of K77, FNS, PDG. However, the diurnal changes of heat storage

within the canopy can be significant for the case of tropical rainforest (*Lindroth et al.*, 2010; *Moore and Fisch*, 1986) due to the high amount of biomass.

Leuning et al. (2012) pointed that energy balance could be closed at the daily scale because when the heat storage energy is averaged over 24 hours, it is negligible compared to the other terms in the energy conservation statement. However, sufficiently large energy closure problems for the study sites remain at the daily (see Figure 4.3) and even the monthly scales (not shown here), pointing to the need to address the issue of imbalance in the eddy flux measurements. Additionally, a relatively lower correlation between R_n and $\lambda E + H$ at K77, BAN, and PDG (e.g., $R^2 < 0.5$ at daily scale, and $R^2 < 0.75$ at hourly scale) may result from significant ground heat flux or data quality issues. This large discrepancy between R_n and $\lambda E + H$ will introduce high uncertainty in the MEP model since Eq (4.16) is used. Therefore, the results of K77, BAN, and PDG are not presented in the result section.

4.4.2 Biomass and air heat storage

An analysis of the heat storage term carried out below further confirms that the energy imbalance mainly results from the measurement uncertainty of eddy fluxes. Specifically, approximation of formulation for heat storage in a vegetation layer proposed in *Moore and Fisch* (1986) and the parameters used in *Ivanov et al.* (2012) are adopted here: the specific heat capacity of wood is 2.93 $[J \cdot g^{-1} \cdot K^{-1}]$, the heat capacity of foliage is 3.50 $[J \cdot g^{-1} \cdot K^{-1}]$, 80% of leaf mass is assumed to be water. The canopy biomass density is estimated from the method defined in *Ivanov et al.* (2012), and a constant wood biomass density is assumed to be 30 $[kg\ C \cdot m^{-2}]$ including both live and dead tissues (*Saatchi et al.*, 2007).

Hourly tower data for the K67 site between 11-16-2017 and 12-11-2017 are selected due to their continuity. The average observed diurnal cycles for R_n, λE_{OBS}, H_{OBS}, and estimated heat storage (S) for the selected time window are shown in Figure 4.4a.

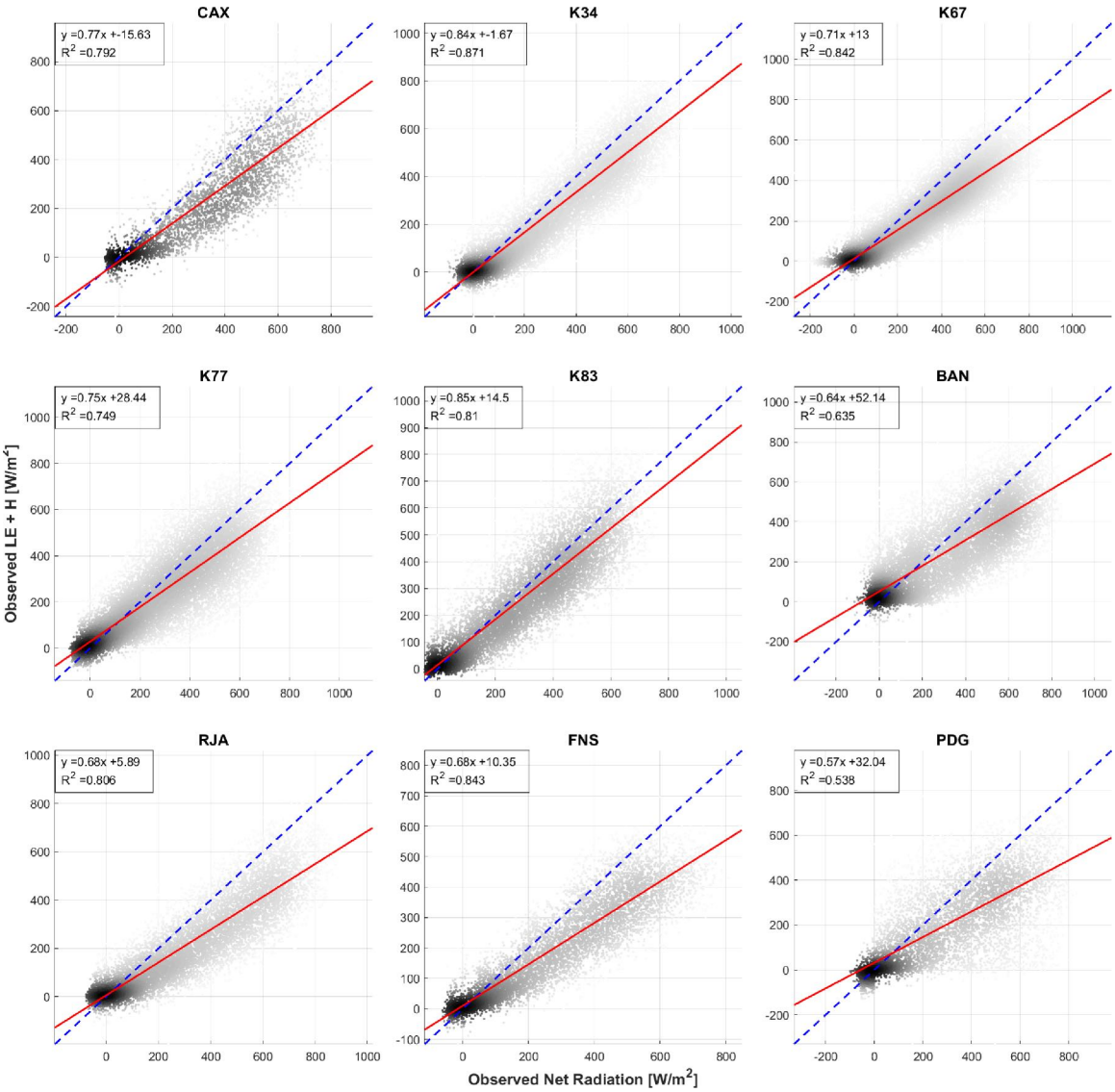

Figure 4.2: Scatter plots of hourly observed net radiation (X-axis) vs. hourly observed latent heat flux plus sensible heat flux (Y-axis) for all of the study sites. A comparison only for the daytime hours (6:00am 19:00pm) is presented here. The blue dashed line is the reference1:1 relationship. The red solid line represents the linear regression line, with the regression equation and R^2 shown in each subplot.

Figure 4.4b is the corresponding scatter plot for the hourly time series comparison. Although this estimation is based on an ad-hoc selection of parameters from prior studies, the result is not overly sensitive to the parameter values within their plausible ranges. Even after adding the estimated heat storage term S to the eddy fluxes $\lambda E_{OBS} + H_{OBS}$, the total heat maintains a significant deficiency in the energy balance

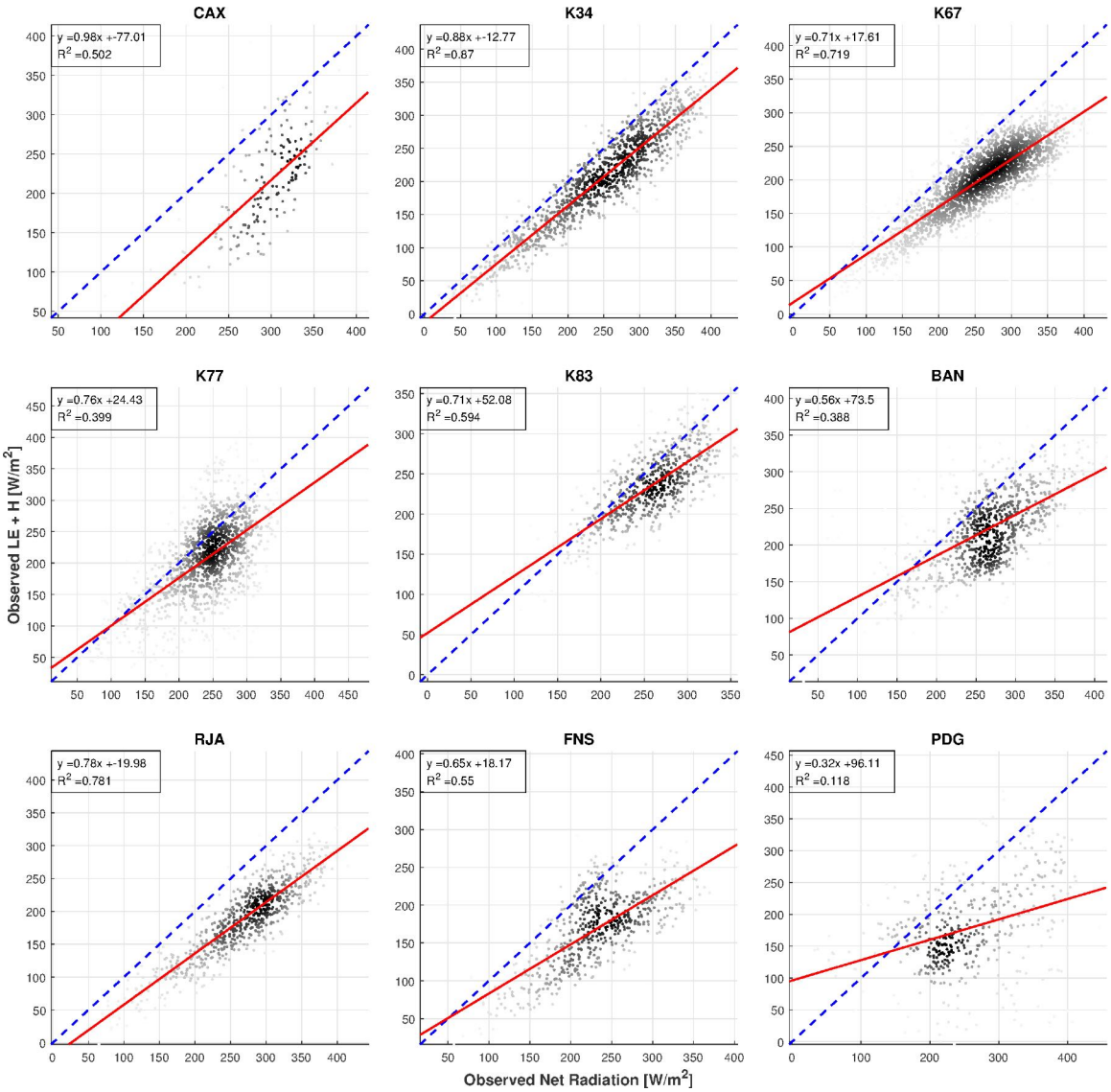

Figure 4.3: Scatter plots of daily observed net radiation (X-axis) vs. daily observed latent heat flux plus sensible heat flux (Y-axis) for all of the study sites. Daily value corresponds to average for daylight hours (6:00am 19:00pm). The blue dashed line is the reference 1:1 relationship. The red solid line represents the linear regression line, with the regression equation and R^2 shown in each subplot.

(Figure 4.4a: $\lambda E_{OBS}+H_{OBS}+S$). As shown in Figure 4.4c, at the daily scale, the heat storage term is negligible, as compared to the other terms in the budget equation. Therefore, the energy imbalance present in the observed series is certainly caused by the measurement errors in λE_{OBS} and H_{OBS} at the daily and monthly scales, as ground heat flux and heat storage components are negligible at these scales.

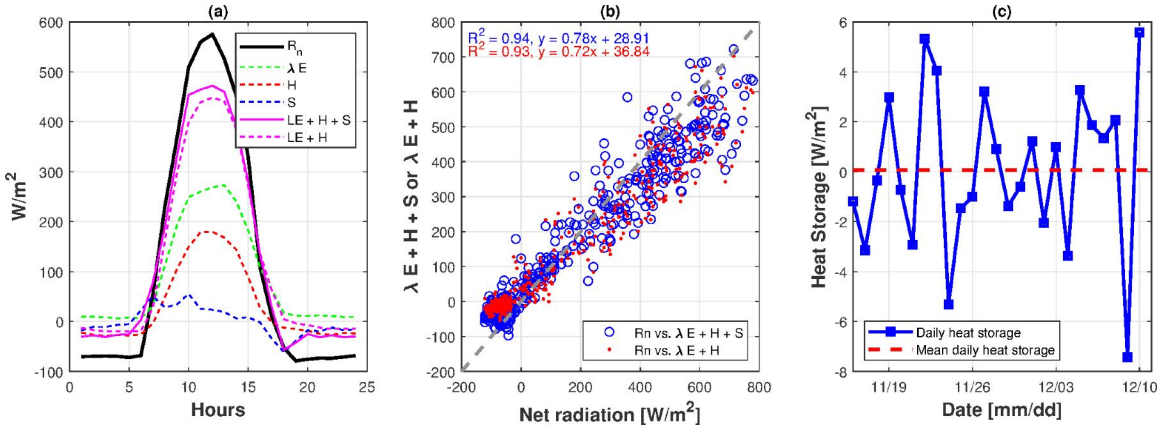

Figure 4.4: Approximated estimation of heat storage by biomass and air at K67. Subplot (a) shows the diurnal cycles of observed net radiation (R_n), latent heat flux (λE_{OBS}), sensible heat flux (H_{OBS}), and heat storage (S). Subplot (b) is the scatter plot for the comparison of R_n and $\lambda E_{OBS} + H_{OBS} + S$ (or $\lambda E_{OBS} + H_{OBS}$), and subplot (c) is the corresponding daily change of heat storage and its mean value over the analyzed period.

4.4.3 Sensitivity analysis of Bowen ratio closure method

We selected K67 for testing the sensitivity of the Bowen ratio closure method by perturbing the error term added to the observed heat fluxes λE_{OBS} and H_{OBS}:

$$\begin{aligned} \tilde{\lambda E} &= \lambda E_{OBS} + \epsilon_{\lambda E} = (1+\alpha)\lambda E_{OBS} \\ \tilde{H} &= H_{OBS} + \epsilon_H = (1+\beta) H_{OBS} \\ R_n &= \tilde{\lambda E} + \tilde{H} \end{aligned} \tag{4.17}$$

where $\epsilon_{\lambda E}$ and ϵ_H represent the error terms added to the original observations to close the energy balance and obtain the corrected fluxes $\tilde{\lambda E}$ and $\tilde{H}$. The parameters α and β represent the ratios of the error term for corresponding observed fluxes. The Bowen ratio closure method in Eq (4.16) corresponds to the case of $\frac{\alpha}{\beta} = 1$, as the ratio between the observed and the corrected fluxes is kept constant. In the sensitivity test, we use different magnitudes of the ratio $\frac{\alpha}{\beta}$ to redistribute the error terms in the corrected fluxes $\tilde{\lambda E}$ and $\tilde{H}$ and subsequently carry out linear least squares regression for the MEP-estimated hourly latent heat and the corrected observation

$\tilde{\lambda E}$. The relationships between the slope and R^2 of the regression line and $\frac{\alpha}{\beta}$ are shown in Figure 4.5 (left panel). The ratio $\frac{\alpha}{\beta} \to 0$ implies adding the residual heat (e.g., $R_n - \lambda E - H$) to $\tilde{H}$, while as $\frac{\alpha}{\beta} \to \infty$ represents the case when the imbalance is biased towards $\tilde{\lambda E}$. *Gerken et al.* (2018) suggested that H may dominate the residual energy between R_n and $\lambda E + H$, implying that the Bowen ratio closure method may not be appropriate in the form of Eq (4.16). Figure 4.5 (right panel) shows that the averaged Bowen ratio is unreasonably high when the residual energy is increasingly redistributed to sensible heat flux. Concurrently, Figure 4.5 (left panel) demonstrates that the corrected flux $\tilde{\lambda E}$ becomes increasingly dissimilar from MEP-based estimates for progressively smaller values of $\frac{\alpha}{\beta}$. Additionally, the measurement errors $\epsilon_{\lambda E}$ and ϵ_H should be of the same order of magnitude, suggesting $\frac{\alpha}{\beta}$ should not close to that two limits (0 or ∞). In summary, the Bowen ratio closure method appears to be an acceptable method to resolve the energy imbalance issue with the ratio $\frac{\alpha}{\beta}$ being close to 1 (Figure 4.5, left panel). The clear advantage is that the method provides the corrected fluxes $\tilde{\lambda E}$ and $\tilde{H}$ that correspond to the energy balance closure.

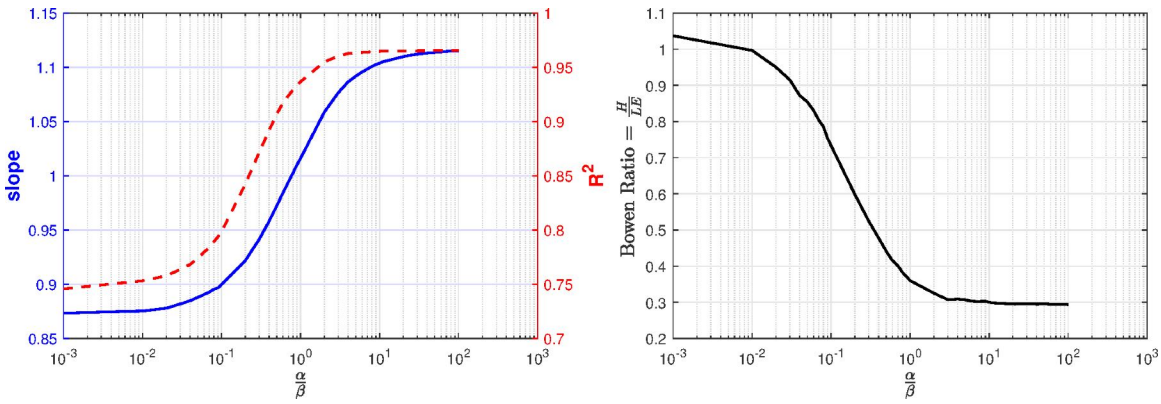

Figure 4.5: Sensitivity analysis of the Bowen ratio closure method using K67 data. The subplot on the left shows the relationship of the MEP performance metrics (evaluated using the least squares regression line slope and the corresponding R^2) with respect to the hypothesized ratio of the flux errors α/β. The subplot on the right shows the averaged Bowen ratio (at the hourly scale) for the corrected $\tilde{\lambda E}$ and $\tilde{H}$.

4.5 Estimation of ET using MEP in Amazon

4.5.1 Validation of MODIS ET for tropical rainforest

Previous studies validated the MOD16A2 Collection 5 product against in situ flux data across a wide range of sites around the world with both satisfactory (*Miranda*, 2017; *Mu et al.*, 2011; *Velpuri et al.*, 2013) and less than adequate performance (*Hu et al.*, 2015; *Long et al.*, 2014; *Wang et al.*, 2015). We report here that for tropical sites across the Amazon basin, the skill of the MODIS ET remains a challenge. Specifically, Figure 4.6(a)-(b) illustrate a comparison of MOD16A2 ET product with observed data at the K67 site (chosen as an exemplary location because of data longevity and reliability). The phase shift, biases, and the low coefficient of determination (R^2) of the MODIS ET product are apparent as shown in the figure, and a similar skill for the other sites can be reported (not shown). Figure 4.6(c) illustrates a comparison using data from all flux sites in the Amazon at the monthly scale: as seen, MODIS ET tends to underestimate at the lower range of evapotranspiration magnitudes and overestimate at the higher range. While inaccuracies may result from biases in the MERRA forcing data, the skill of MODA216 is not improved when estimation relies on site-level observations, as reported by their user guide (`https://lpdaac.usgs.gov/products/mod16a2v006/`).

4.5.2 MEP model performance at Amazon rainforest

A comparison between MEP-based estimates and original/uncorrected observations (Figure 4.7) shows that MEP tends to overestimate ET for all sites. There are two possible reasons for that: (a) water stress is not accounted for since the surface resistance parameter η_s is ignored; (b) the energy budget imbalance in the observed data. To determine whether water stress is responsible for the ET estimation using the MEP method, the observed λE is first corrected using the Bowen ratio closure

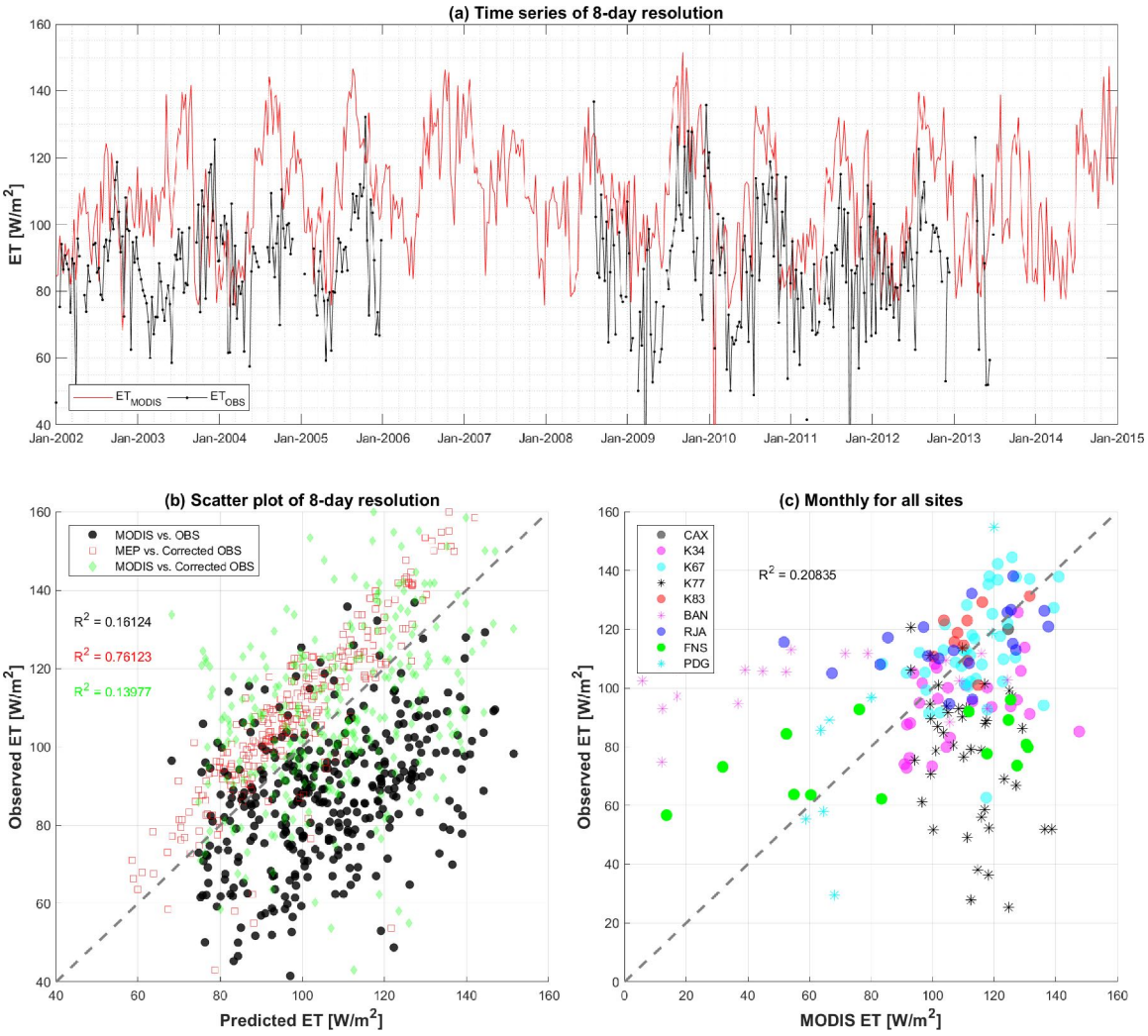

Figure 4.6: A comparison of the MODIS ET product (denoted as MODIS) for the K67 site. Subplot **(a)** shows a comparison of the time series at 8-day resolution (i.e., average over 8 days): OBS refers to the data described in *Restrepo-Coupe et al.* (2013) and Corrected refers to flux magnitudes adjusted using Eq (4.16). Subplot **(b)** compares fluxes in **(a)** in the form of a linear regression, also adding MEP-modeled ET resolved at the same temporal scale. MODIS ET is compared with the corrected monthly ET observations for 9 sites across the Amazon in plot **(c)**. R^2 was calculated excluding data for K77, BAN, and PDG whose observed surface fluxes were intermittent. The grey dashed lines in subplot **(b)**, and **(c)** represents the reference 1:1 relationship.

method as in Eq (4.16) to satisfy energy conservation.

Figure 4.8 illustrates a close agreement between the MEP modeled and the corrected observed ET at hourly scale with the regression exhibiting high R^2 (about 0.9) and the slope of the relationships being close to 1 for the majority of the sites. The results for K77, BAN, and PDG sites are not shown due to data quality issues. Although the simulated ET tends to be somewhat lower than the corrected observations overall, it is consistent with the replacement of q_s with q_a. Specifically, q_a should underestimate q_s when λE is positive (e.g., away from the vegetated surface to the atmosphere), leading to lower estimates of the heat flux. Further, using Eq (4.11) with the assumption of $\eta_s = 1$ (i.e., no water stress) as a surrogate of q_s leads to similar MEP performance, but higher magnitudes (not shown here). Using saturated humidity in the MEP theory provides an estimate of potential ET.

4.5.3 Invariance of MEP model skill with temporal scale

While the MEP model performs well at the hourly scale (section 4.5.2, also see *Wang and Bras* (2011) for other environments), in many hydrologic analyses it is sufficient to have daily or even monthly resolution. This is particularly true for areas where accurate hourly data are unavailable, and this is representative of most of the tropics. Figure 4.8g shows that MEP modeled ET is in close agreement with observations at the monthly (and daily, not shown here) scale. Input data at site-level are expected to have higher accuracy than reanalysis products that are also representative of coarser scales. Nonetheless, using the SYN-1deg monthly CERES net radiation and surface temperature data (Figure 4.8h) also results in a closer agreement of MEP model estimates with observations as compared to MOD16A2 ET (Figure 4.6c).

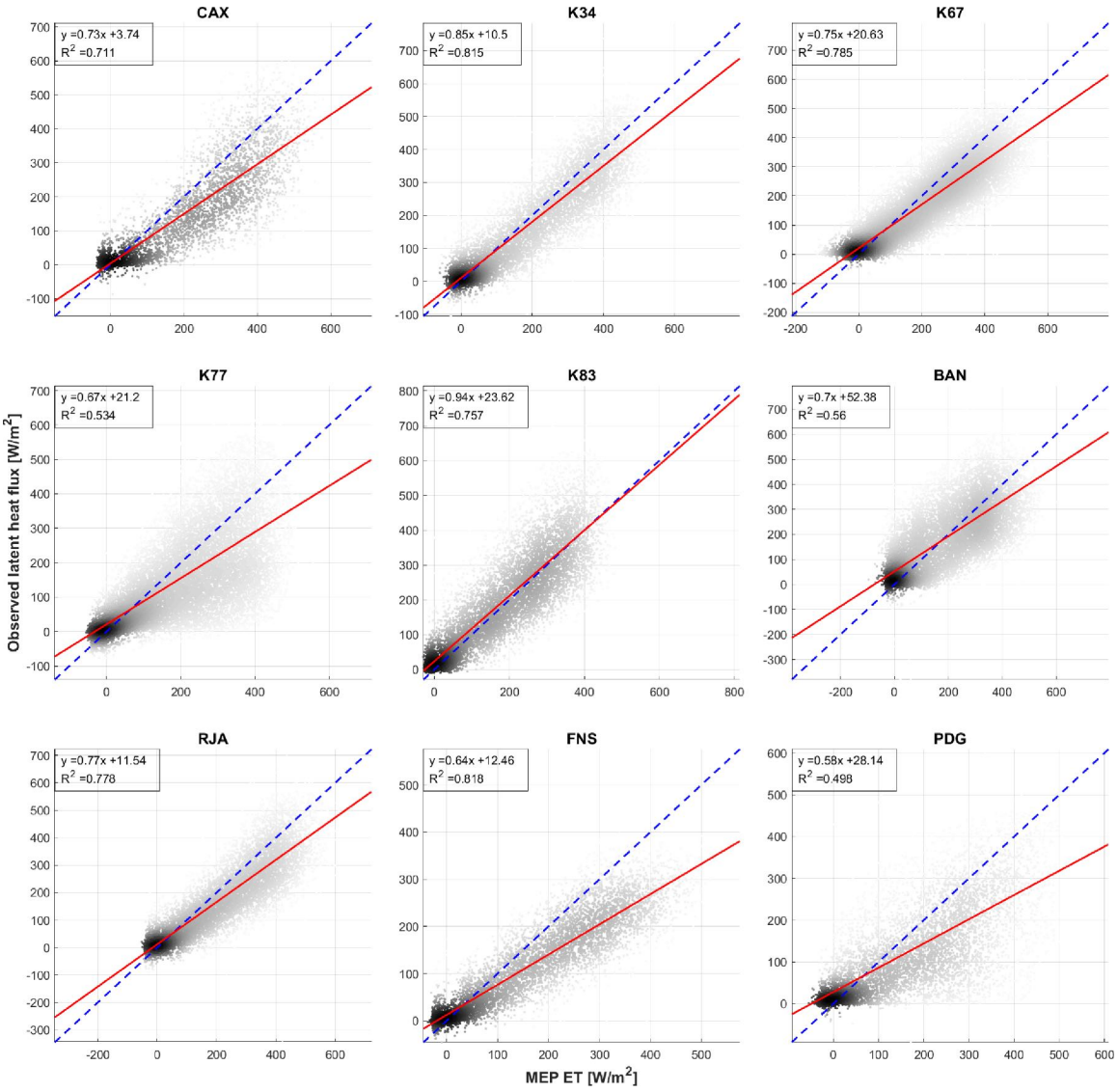

Figure 4.7: Scatter plots of hourly MEP simulated latent heat flux (X-axis) vs. observed (uncorrected) latent heat flux (Y-axis) for all of the study sites. A comparison only for the daytime hours (6:00am 19:00pm) is presented here. The blue dashed line is the reference 1:1 relationship. The red solid line represents the linear least-squares regression line, with the equation and R^2 shown in each subplot.

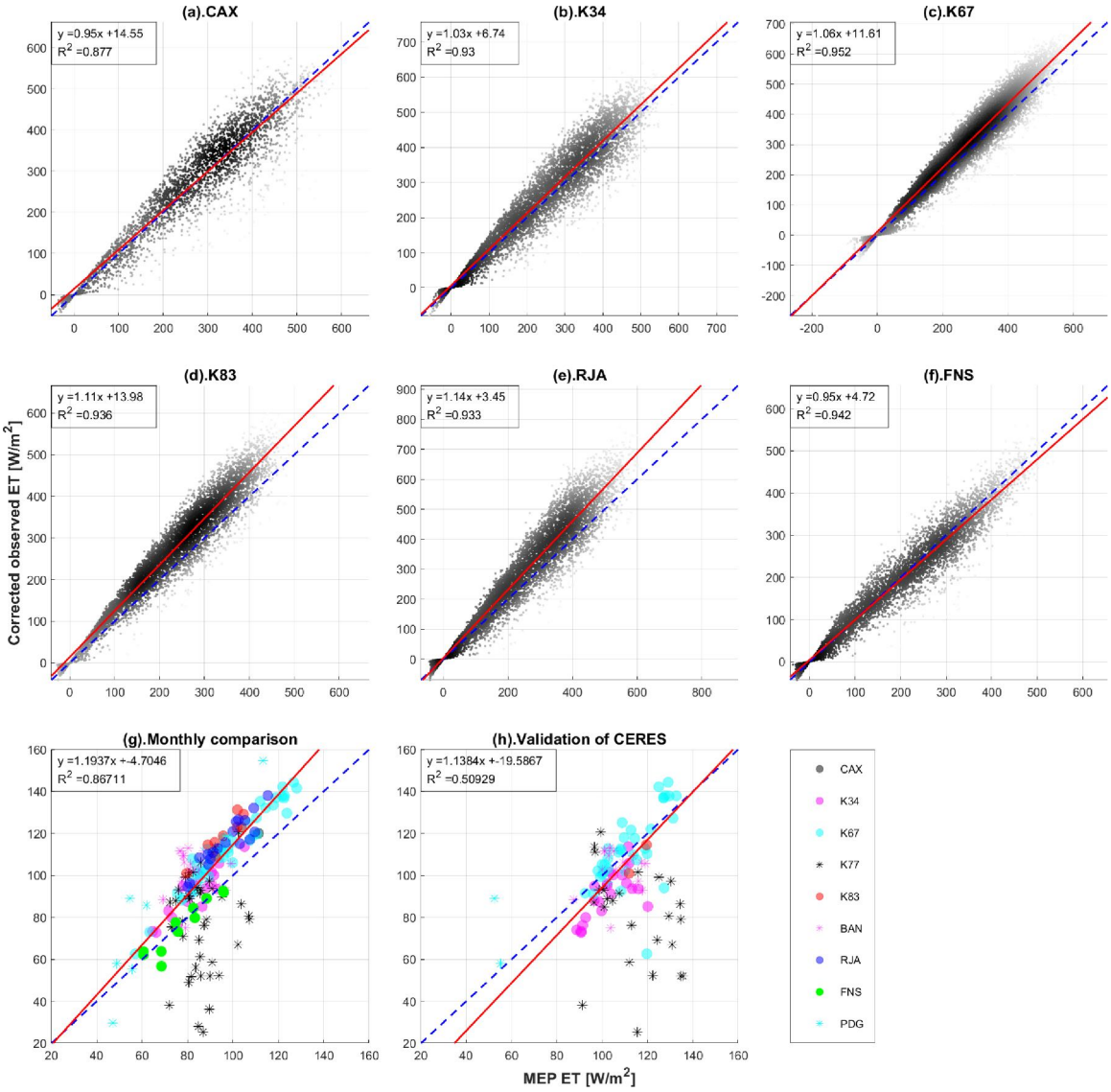

Figure 4.8: Scatter plots of MEP estimated ET (X-axis) against observed ET (Y-axis) corrected using the Bowen ratio closure method (Eq (4.16)) for selected study sites at hourly scale (subplot **(a)** to subplot **(f)**) and at monthly scale (subplot **(g)** and subplot **(h)**). Comparison only for the daytime hours (6:00am 19:00pm) at the hourly scale is presented. In subplot **(g)**, MEP is driven by observed net radiation, air temperature, and air specific humidity. In subplot **(h)**, MEP is driven by CERES net radiation and surface temperature. The blue dashed line is the reference 1:1 relationship. The red solid line represents the linear least-squares regression line, with the equation and R^2 shown in each subplot (at monthly scale, the data for K77, BAN, and PDG are excluded).

4.5.4 MEP ET under drought conditions

Although MEP model estimates agree well with observations without considering drought impacts, vegetation control of water loss can be an important factor during dry intervals of high-intensity (*Costa et al.*, 2010). Soil matric potential in the root zone drives plant response, but relevant data are never available, and suitable proxies are typically used. To assess MEP performance at varying moisture conditions, we adopted vapor pressure deficit (VPD) and cumulative water deficit (CWD) to represent drought severity, and ϵ_{ET} is defined below to denote the MEP performance:

$$\epsilon_{\lambda E} = \lambda E_{MEP} - \lambda E_{corrected} \tag{4.18}$$

where $\lambda E_{corrected}$ is computed from Eq (4.16). Drier conditions are associated with higher VPD, and therefore, ET estimates without explicit inclusion of drought control on the function of plant stomata are expected to be higher than measured ET. However, there is no significant relationship between VPD and $\epsilon_{\lambda E}$(Figure 4.9a), indicating the MEP model performs consistently across a range of VPD. Likewise, there is no significant relationship found between $\epsilon_{\lambda E}$ and CWD (Figure 4.9b). Chi-Square test of independence (*McHugh*, 2013) is used to test the relationship between the performance and drought indicators with the null hypothesis that MEP performance is independent with droughts. The p-value for pairs of data sets in subplots (a) and (b) of Figure 4.9 are 0.2391, and 0.2928 respectively. Thus, we can accept the null hypothesis that the performance of the MEP model is not affected by any drought level in our validate data.

4.5.5 ET estimation at basin scale

There is a substantial difference in the spatial distribution of ET between MEP and MODIS derived estimates (see Figure 4.10 as an illustration), which are driven

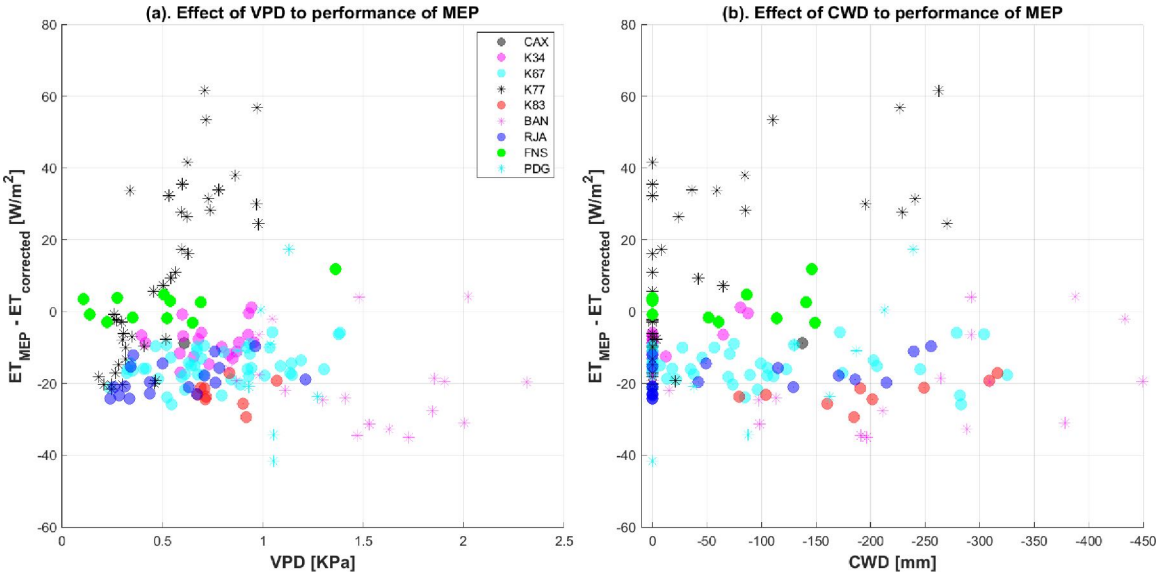

Figure 4.9: The effects of **(a)**. VPD, and **(b)**. CWD on the performance of MEP-based estimates ($\lambda E_{MEP} - \lambda E_{corrected}$).

by different data products (CERES vs. MERRA). To address the comparison uncertainty caused by differences in inputs, we run MEP analysis with: (1) CERES net radiation and surface temperature; (2) CERES net radiation and MERRA air temperature and air specific humidity at 2 m; and (3) MERRA net radiation and CERES surface temperature. The results are scaled up by averaging ET over the Amazon basin. Figure 4.11 illustrates a comparison of these scenarios. The original MODIS ET product has a prominent phase shift when compared to the MEP ET that uses CERES data (case 1). From the similarity of the time series for cases 1 and 3, one may conclude that net radiation in MERRA and CERES is in close agreement (but note the peak differences in Figure 4.11b). The MEP estimates for case 2 are quite lower, dropping below MODIS ET magnitudes, implying that the source of specific humidity is quite important.

Although the monthly MEP ET (driven by CERES data) and MODIS ET (driven by MERRA) have discrepancies at the basin scale, the long-term annual ET is consistent between the two estimates. For example, the annual ET averaged from 2003 to 2013 is **1,280 mm** (MEP ET) and **1,284 mm** (MODIS ET). Both methods over-

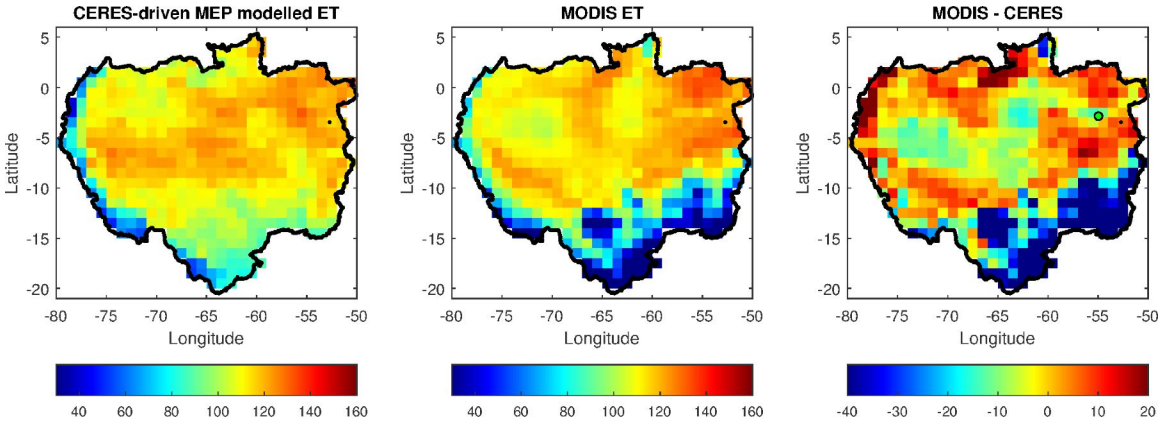

Figure 4.10: ET estimates from the MEP theory (left panel, driven with CERES net radiation and surface temperature) and MODIS ET product (central panel: driven with MERRA meteorological data) for August 2008. Average daily latent heat flux is illustrated; the unit in the color bar is $[W/m^2]$. The original resolution of MODIS (1000 m×1000 m) is upscaled to the CERES resolution (1 deg×1 deg) by averaging values in MODIS grid cells in the corresponding CERES grid cell. The difference between the two products (MODIS - MEP-based estimates) is illustrated in the right panel.

estimate ET, as compared to the water budget method, which yields **1,088 mm**, but is still within the range suggested in earlier studies (*Maeda et al.*, 2017).

4.5.6 Discussion

Our results show that the performance of the MEP model has an adequate skill in simulating ET across the Amazon. The model performance is consistent under a variety of possible water stress situations, even though drought impacts on water flux are not included explicitly. The two indicators representing the drought intensity show no clear, statistically significant signal in the differences between the MEP ET and observed ET. This fining provides indirect evidence in support of previous research, arguing that the Amazon rainforest is resilient to seasonal water limitations and mild droughts (*Davidson et al.*, 2012). Another possibility is that the effect of canopy constraint on water flux is already reflected in MEP inputs net radiation or air specific humidity. Further evaluation is necessary under the conditions of extreme water limitation, such as the 2015 drought (*Jiménez-Muñoz et al.*, 2016).

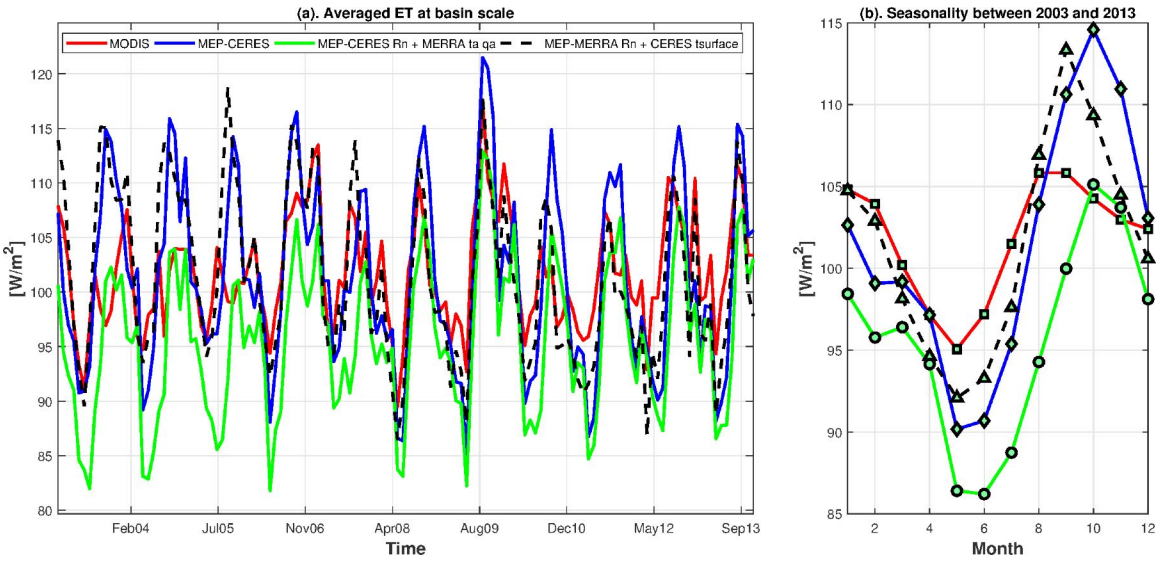

Figure 4.11: A comparison of ET between MODIS and MEP based estimates at the scale of the Amazon basin. The red solid line represents the monthly MODIS ET product downloaded from the NTSG website; the other times series represent MEP analysis carried out using inputs from different meteorological data sets: (1) net radiation and surface temperature from CERES (the blue solid line); (2) net radiation from CERES, while air temperature and air specific humidity at 2 m from MERRA (the green solid line); and (3) net radiation from MERRA and surface temperature from CERES (the black dashed line). The corresponding seasonality is presented in subplot **(b)**.

Site-level MEP ET is more consistent as compared to the MOD16A2 product. The inferior performance of the MODIS product is likely due to the high biodiversity of the Amazon rainforest, making it difficult to characterize the resistance terms in Eq (4.2). Specifically, the MOD16A2 product uses LAI to parameterize canopy processes and resistance (*Mu et al.*, 2011), which is highly variable and challenging to capture with remote sensing (*Aragão et al.*, 2005; *Hilker et al.*, 2015).

Both MEP ET and MODIS ET yield similar long-term averages at the Amazon basin scale but are higher than the average obtained with the water budget analysis. This overestimation with the MEP model might be caused by lack of inclusion of water limitation effect during droughts: while we do not find a clear signature of droughts at the site-level comparison, the two major droughts in the analysis period (i.e., 2005 and 2010) may not have been pronounced at the flux sites with observational data.

But other areas in the basin can be affected by the drought significantly. Using the saturated specific humidity q^{sat} from Eq (4.11) as a proxy for q_s (i.e., $\eta_s = 1$) expectedly results in ET overestimation. Using q_a as an alternative proxy for q_s might address the drought effects. For example, the MEP ET corresponding to case 3 in Figure 4.11b yields basin-scale ET of **1,200 mm**, which better matches the estimate from the water budget analysis. Another possible contributing factor is the overestimation of net radiation by the CERES product, as shown in Figure 4.12, also previously reported for the continental U.S. by *Jin et al.* (2011).

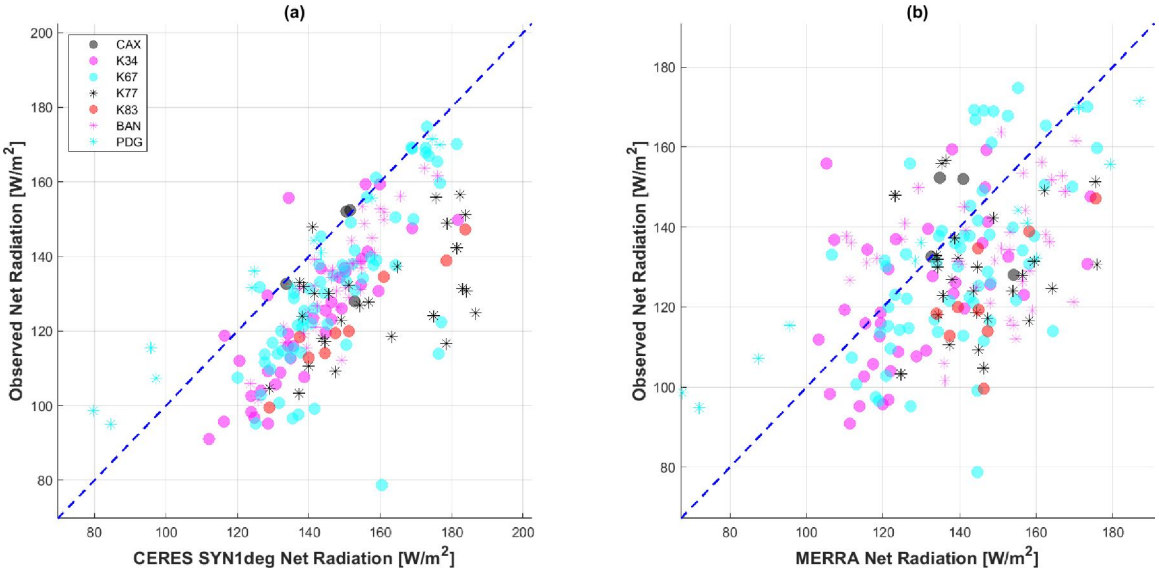

Figure 4.12: Validation of **(a)**. CERES, and **(b)** MERRA net radiation products with tower observation. Y-axis represents observed monthly net radiation in both subplots. X-axis represents the CERES and MERRA monthly net radiation respectively.

Furthermore, higher MEP ET is obtained in the basin southeast (Figure 4.10), characterized by savannas, grasslands, and croplands (∼17.5% of the basin, Figure 4.13). It is reasonable to assume that ground heat flux is non-negligible for these land-surface types. But this heat flux is ignored here, leading to higher ET from non-forested areas, e.g., **1115 mm** (MEP) vs. **979 mm** (MODIS) in 2008, as an example. Hence, the MEP ET at the basin scale can be further improved by addressing ground heat flux in sparsely vegetated areas.

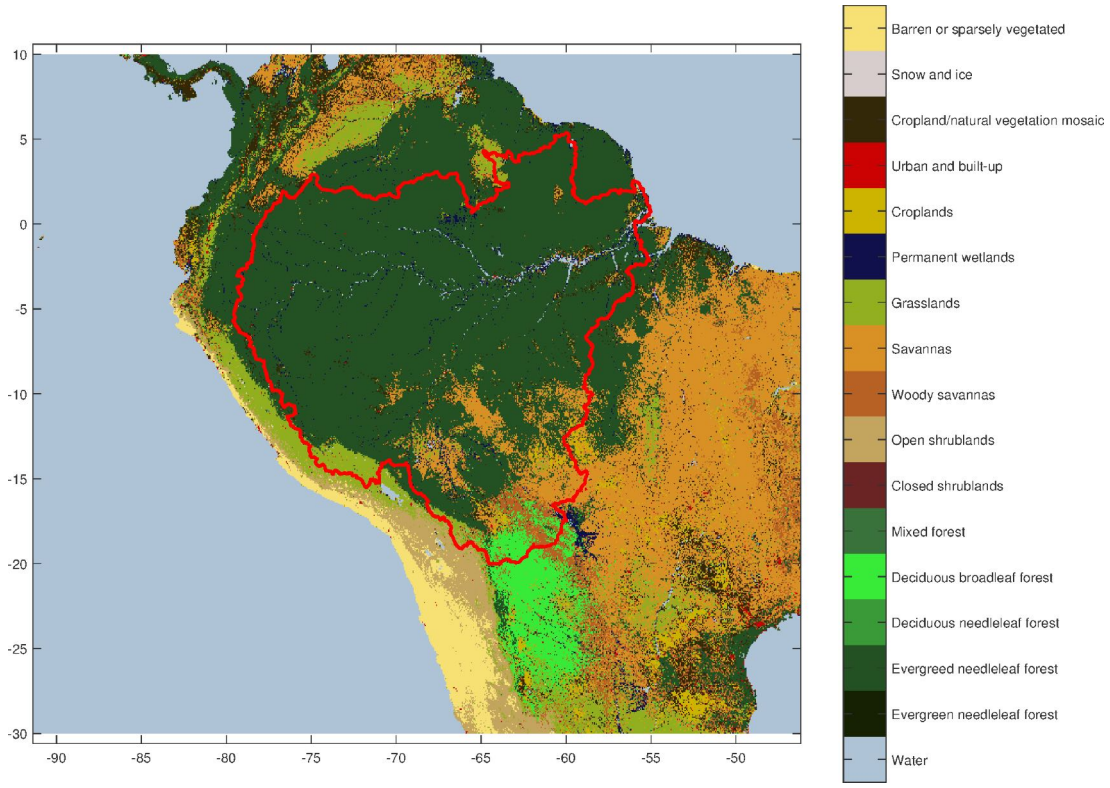

Figure 4.13: Land cover type for 2008 from MODIS Land Cover Product (MCD12Q1 Type-1). The classification scheme associated with each color is listed in the legend. The red line is the watershed boundary corresponding to Obidos station, which is used for basin scale analysis in this chapter.

While the MEP model is parsimonious and does not require variables of high uncertainty (e.g., LAI, surface resistance, wind speed, roughness lengths, etc.), a relevant question is whether model inputs are readily available. Fairly accurate remote sensing (e.g., CERES, Figure 4.12) and reanalysis data (e.g., MERRA) of net radiation and temperature are becoming increasingly available. Surface humidity data are however unavailable, but can be estimated from soil moisture and temperature (*Hajji et al.*, 2018) or replaced by near-surface air humidity. Thermal inertia of soil is another MEP parameter for bare soil (not used here), which can be parameterized as an empirical function of soil moisture through statistical analysis. The increasing abundance and quality of data sources support the possibility of future regional and global ET assessments with the MEP model.

Overall, the presented analysis demonstrates the ability of the MEP model to simulate ET in the Amazon rainforest realistically. We underline that the model performance can be attributed to efficient use of available information about four aspects of the ET process: energy supply, liquid water supply, evaporation potential, and transport of water vapor (*Shuttleworth*, 1993). The theoretical foundation of the MEP method is the well-established maximum entropy principle in the Bayesian probability theory and information theory (*Jaynes*, 2003). The information entropy guarantees the relevance and logical consistency of the information provided by the input variables (net radiation, surface temperature, and humidity) to ET parameterization. Additionally, the extremum solution of the Monin-Obukhov similarity equations (*Wang and Bras*, 2010) allows one to avoid using gradients in the MEP model formulation. In summary, it is an attractive method to assess regional energy budgets in tropical rainforest environments. Accurate ET estimates from MEP may help understand mechanisms of transition from dry to wet season (*Wright et al.*, 2017), investigate the legacy effects of droughts (*Saatchi et al.*, 2013), and improve precipitation modeling (*Yin et al.*, 2013) over the Amazon.

4.6 Climate change impact on ET in future

Outputs from Global Climate Models (GCMs) are usually used to study the impact of climate change on the variable of interest. In this section, we selected GCMs from 12 individual institutions that participate in the CMIP5 based on the availability of required inputs, i.e., net radiation, specific humidity, near-surface temperature, leaf area index, and top layer soil moisture (Table 4.2). These models also provide outputs of latent heat flux (e.g., ET), which will be used for comparison with MEP-based ET estimates. Three periods are selected: 1961-1990 for the control period (CTL), 2041-2070 for mid-century (MID), and 2071-2100 for the end-century (END). RCP45 and RCP85 are used in this chapter to represent the emission scenario in the future period (FUT). Outputs from all the GCMs are converted to $1^\circ \times 1^\circ$ with bi-linear interpolation for computational convenience.

Table 4.2: A list of CMIP5 models used for evaluating ET in MEP

No.	Institution	Model name	Resolutions	References
1	BNU	BNU-ESM	128 × 64	*Wei et al.* (2012)
2	CCCMA	CanESM2	128 × 64	*Chylek et al.* (2011)
3	CSIRO_BOM	ACCESS1-0	192 × 145	*Bi et al.* (2013)
4	INM	Inmcm4	180 × 120	*Volodin et al.* (2010)
5	IPSL	IPSL-CM5A-LR	96 × 96	*Dufresne et al.* (2013)
6	MIROC	MIROC5	256 × 128	*Watanabe et al.* (2010)
7	MOHC	HadGEM2-ES	192 × 145	*Jones et al.* (2011)
8	MRI	MRI-CGCM3	320 × 160	*Yukimoto et al.* (2012)
9	NCAR	CCSM4	288 × 192	*Gent et al.* (2011)
10	NCC	NorESM1-M	144 × 96	*Bentsen et al.* (2013)
11	NOAA_GFDL	GFDL-ESM2G	144 × 96	*Dunne et al.* (2012)
12	NSF_DOE_NCAR	CESM1-CAM5	288 × 192	*Meehl et al.* (2013)

4.6.1 Performance of MEP-based ET driven by GCM forcings

Since the observations of ET are not abundant, and the measurement periods are always short, the precipitation and runoff data of *Livneh et al.* (2013) are used in Eq. (4.3) to estimate mean annual ET during the control period (1961-1990) over the

US continent. The performance of the water budget based ET can be guaranteed by validating the precipitation and runoff with WebMET precipitation (see Appendix B) and USGS streamflow data (see Figure 3.3) respectively. Figure 4.14 illustrates the spatial pattern of annual mean ET of control period (e.g., 1961-1990) from *Livneh et al.* (2013), multi-model mean of MEP-based ET, and ET from GCM outputs. ET estimates from each individual GCM can be found in Figure 4.15 and Figure 4.16 for MEP method and original outputs, respectively. Figure 4.17 shows the pixel to pixel scatter plots for each individual GCM in terms of original ET and MEP-based ET. As indicated by the R^2 and slope of the linear regression in Figure 4.17, MEP modeled ET has similar performance as the GCM-based ET but lower magnitudes.

Aggregating the spatial ET to continent scale, *Livneh et al.* (2013), multi-model mean of MEP and GCM yield **555 mm/year**, **522 mm/year**, and **664 mm/year**, respectively. The box plots of the annual ET during the control period at the continent scale for each GCM is illustrated in Figure 4.18. MEP method yields much lower estimates of ET as compared to GCM outputs, even though MEP estimation uses a subset of the same variables involved in ET simulation by any given GCM. Further, *Livneh et al.* (2013) reported that their products tend to overestimate ET, where it is shown that precipitation in *Livneh et al.* (2013) overestimates observed precipitation from the station-based dataset by WebMET and runoff of *Livneh et al.* (2013) underestimates the observed streamflow from the USGS in average. Another data product used in the comparison is from the National Climate Assessment - Land Data Assimilation System (NCA-LDAS; *Jasinski et al.*, 2019), which yields **507 mm/year** for the averaged annual ET between 1979 and 1990 at the US continent. Therefore, the MEP method provides a more consistent annual ET estimates as compared to reference data.

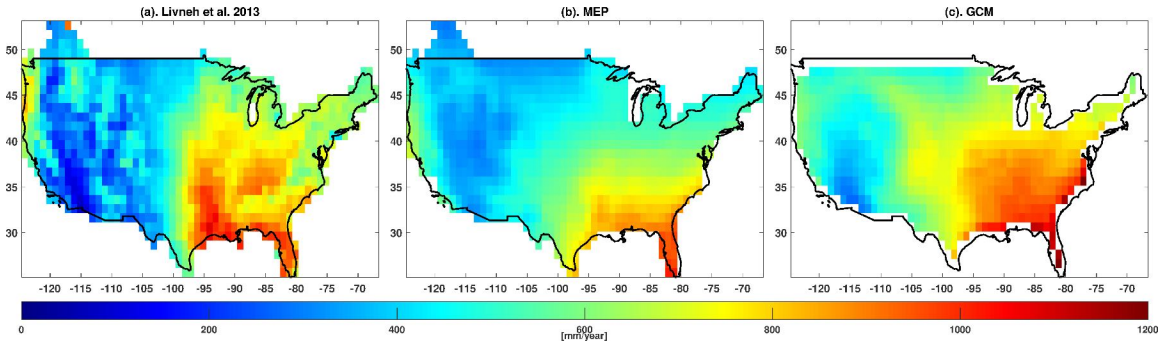

Figure 4.14: Mean annual ET over the US continent for the period of 1961-1990 using (a). *Livneh et al.* (2013) data set to estimate with the water budget method, (b). arithmetic mean of ET from the multi-model ensemble of MEP-based estimates, and (c). arithmetic mean of ET from the multi-model ensemble of GCM-based estimates.

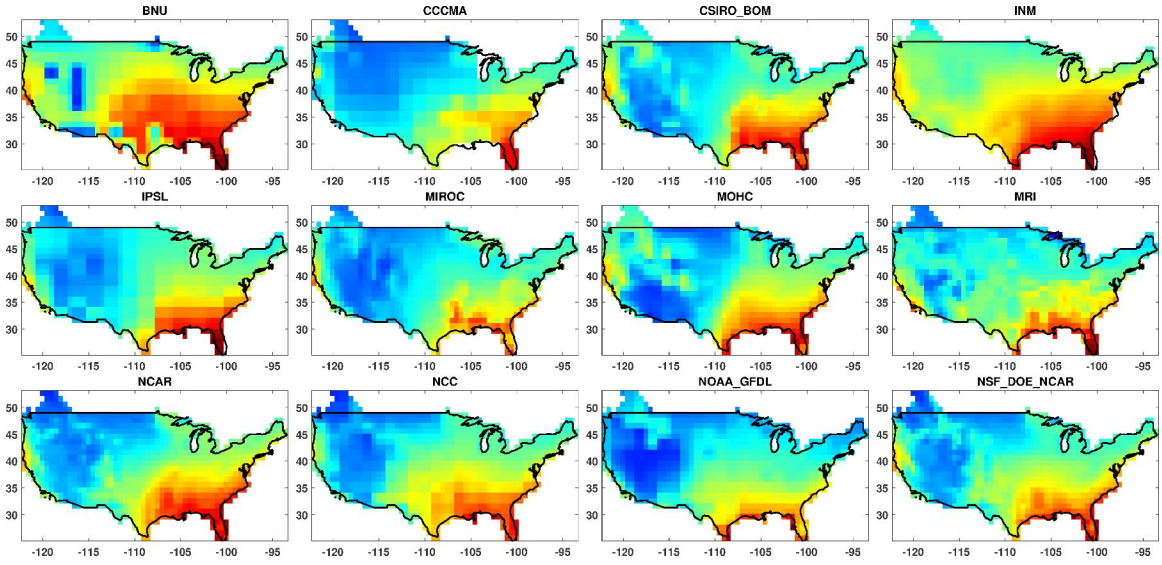

Figure 4.15: Mean annual ET over the US continent averaged from 1961 to 1990 for each individual GCM estimated using the MEP method. The color scale is the same as that of Figure 4.14.

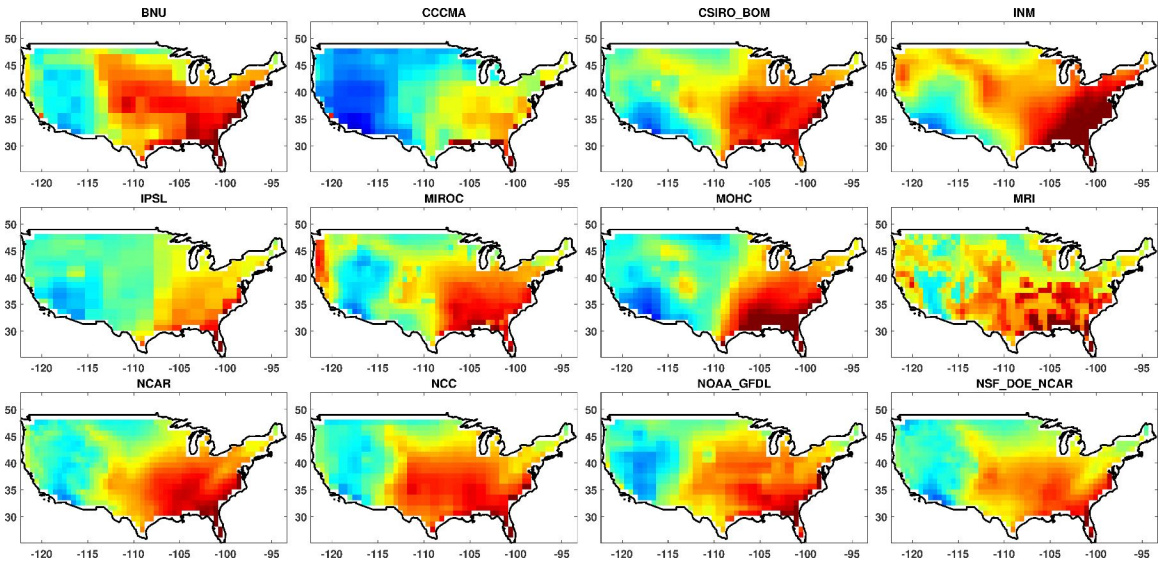

Figure 4.16: Mean annual ET over the US continent averaged from 1961 to 1990 for each individual GCM estimate. The color scale is the same as that of Figure 4.14.

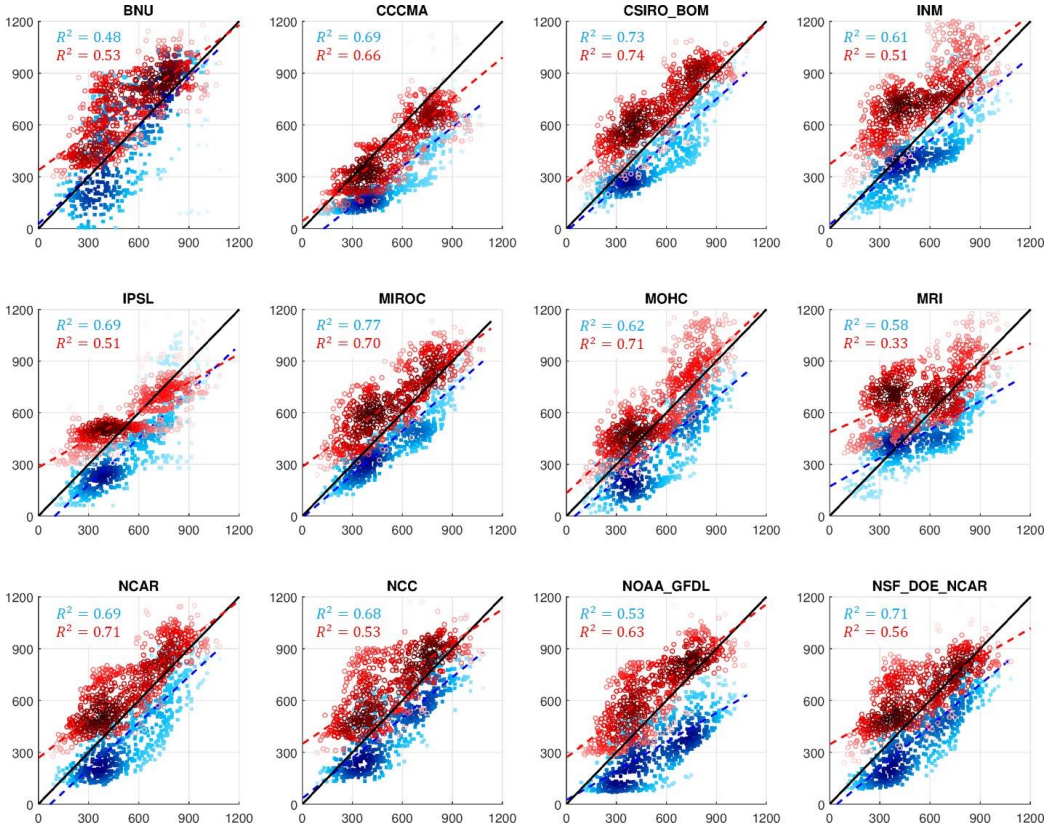

Figure 4.17: Pixel to pixel scatter plot for annual mean ET validation. The x-axis represents the water budget ET estimates from *Livneh et al.* (2013). The y-axis represents the MEP-based ET estimates (blue squares) and GCM original ET (red circles). Darker color is associated with the density of the scatter plot for both.

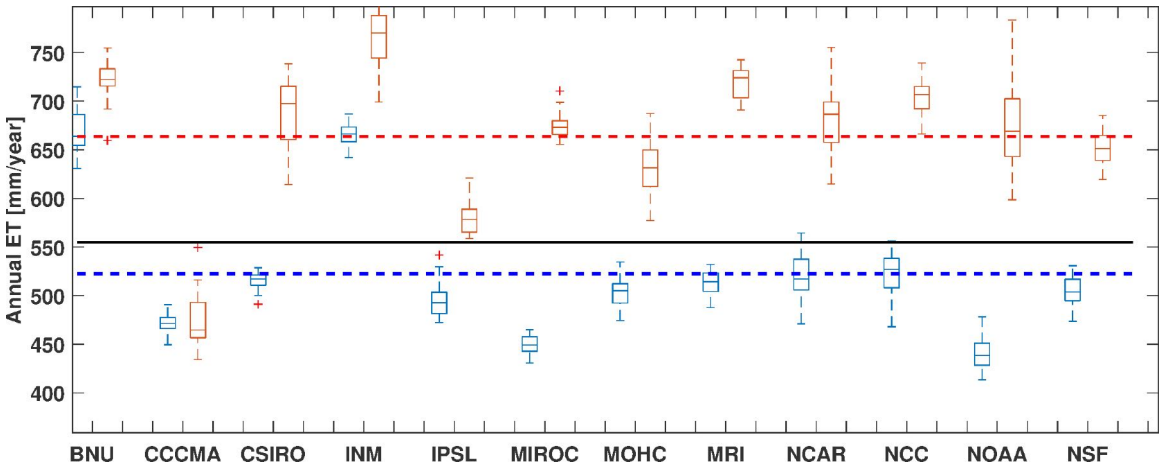

Figure 4.18: Boxplots of annual ET at the continent scale for each individual GCM during control period. Specifically, the red box plots are for the GCM-based ET estimates and the blue box plots are for the MEP-based ET estimates. The black solid line represents the mean value from the *Livneh et al.* (2013) product, the red dashed line denotes the multi-model ensemble mean from GCM-based estimates, and the blue dashed line is the multi-model ensemble from MEP-based estimates.

4.6.2 Projections of ET from MEP

MEP model is run for two future periods and two emission scenarios, which corresponds to four different cases: a). MID-RCP45; b). MID-RCP85; c). END-RCP45; and d). END-RCP85, where "MID" refers to the mid-century, "END" refers to the end-century, and RCP45 and RCP85 are the two emission scenarios. As suggested in section 3.3, these four cases correspond to different increases of temperature in the future periods. For example, the averaged temperature over the US continent can increase about $2.5°C$, $3.2°C$, $3.0°C$, and $5.0°C$ under future scenarios of MID-RCP 45, MID-RCP 85, END-RCP 45, and END-RCP 85, respectively. Projections of ET change in the future can be found in Figure 4.19 for MEP-based estimates, and Figure 4.20 for GCM-based estimates. Representing the highest temperature increased in the future, end-century RCP 85 shows the most considerable magnitude of change.

When considering the same future scenario, MEP-based estimates exhibit a lower increase of ET than ET simulated by GCMs. This difference is likely caused by the

inclusion of soil moisture in MEP formations (see Eq (4.8), (4.12)). Soil moisture represents liquid water supply for transpiration through plant functionality, which is a key aspect of the ET process (*Shuttleworth*, 1993). It is reported in *Dai* (2013) that the climate models project that the soil moisture will decrease in most US regions, giving rise to an increased risk of drought. This predicted drought represents an increase of water limitation in soil moisture that can suppress the increase of ET while higher temperatures, humidity, and radiation are projected in the future.

Metric of *Knutti and Sedláček* (2013) is used to represent the robustness of the models' projections. The robustness metric (e.g., R) is defined as the ratio of model spread to the predicted change. If $R = 1$, it means the model noise is negligible compared to their predicted change. Otherwise, it yields a small or negative value (see section 3.2.2). The models show higher robustness for the projected change of ET with the MEP method as compared to GCM-based ET estimates, such as there are more cells with high robustness (e.g., $R > 0.6$) for the former method (see Figure 4.19 and Figure 4.20). Furthermore, the predicted changes from original GCM outputs are more significant for all the four selected cases but exhibit lower robustness than that of MEP projections. This means the model spread or uncertainty of ET projections is reduced by using the MEP method.

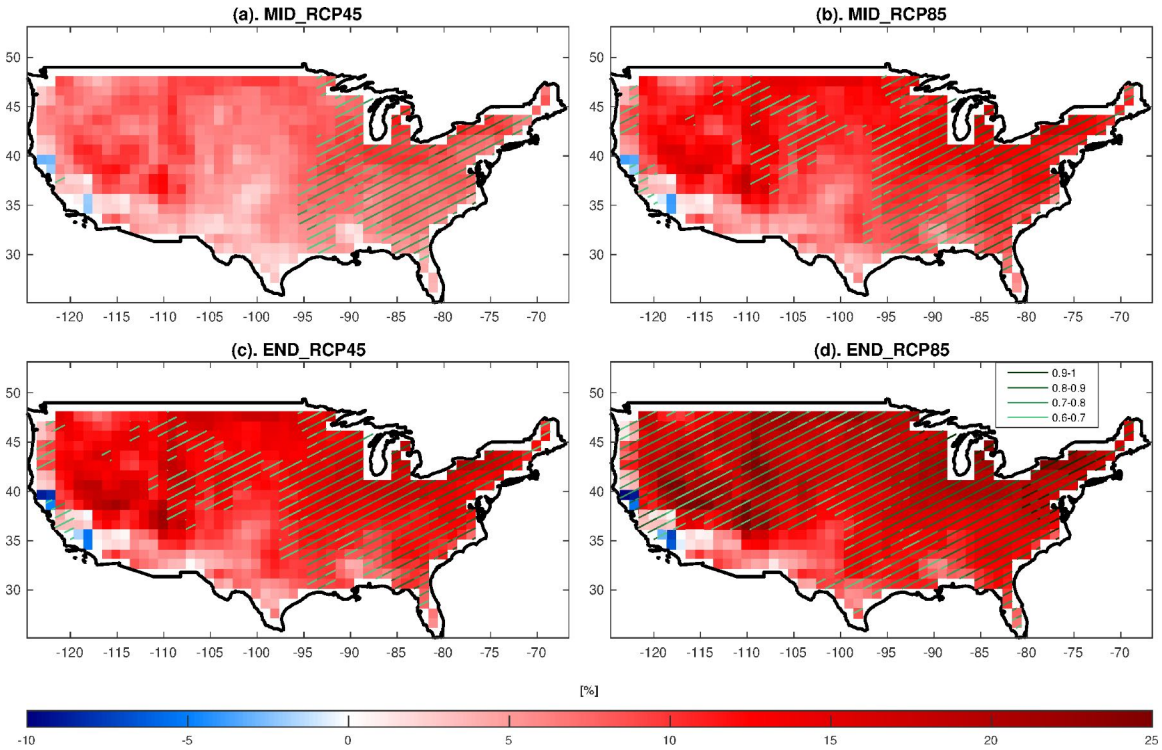

Figure 4.19: Projected change of ET from multi-model mean of MEP based ET for (a). MID century, RCP45; (b) MID century, RCP 85; (c) END century, RCP45; and (d) END century, RCP85. The change is defined as $\frac{Future - Control}{Control} \times 100[\%]$. The robustness of models' projection are evaluated with method of *Knutti and Sedláček* (2013). Grid cells with high robustness (e.g., robustness metric > 0.6) are overlaid with green slash lines. The darker color is associated with high robustness metric.

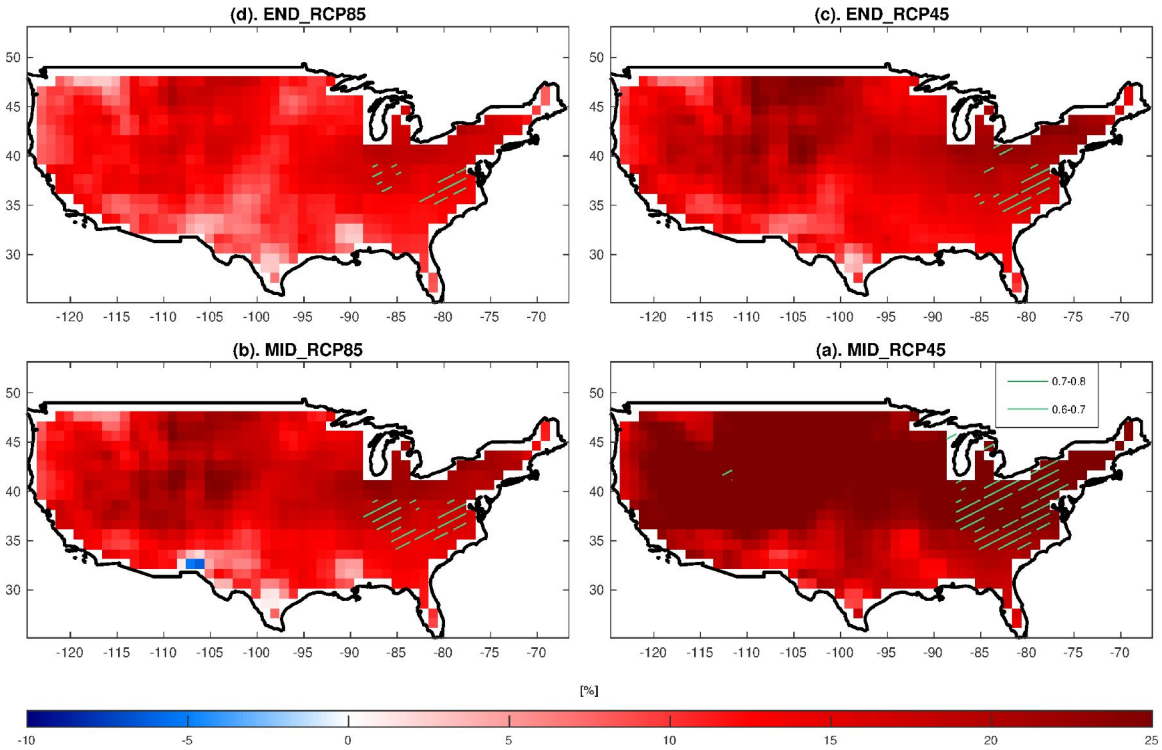

Figure 4.20: Projected change of ET from the multi-model mean of GCM-based ET estimates for (a). MID century, RCP45; (b) MID century, RCP 85; (c) END century, RCP45; and (d) END century, RCP85. The change is defined as $\frac{Future - Control}{Control} \times 100[\%]$. The robustness of models' projection is evaluated with method of *Knutti and Sedláček* (2013). Grid cells with high robustness (e.g., robustness metric > 0.6) are overlaid with green slash lines. The darker color is associated with high robustness metric.

4.6.3 Discussion

MEP method shows better skills in reproducing ET using GCM outputs. Specifically, MEP produces more consistent ET with other ET products than GCMs for the control period at the US continent scale. The multi-model mean of GCM-based ET estimate is about 20% higher than the *Livneh et al.* (2013) estimates, whose quality is approved by checking the water budget using independent data sources: the USGS streamflow and WebMET precipitation. This discrepancy is significant since the estimates are derived from the 30-year annual average at the continent scale. Thus, we need to understand the source of additional water available for ET in the water budget. For the annual precipitation and runoff of the control period averaged at the continent scale, the CMIP5 multi-model mean is **884mm/yr** and **220mm/yr**, respectively. For the *Livneh et al.* (2013) product, these estimates are **774mm/yr** and **226mm/yr**. Additionally, *Mehran et al.* (2014) reported CMIP5 multi-model ensemble overestimates precipitation in the west of the US, which correlates with the overestimation of ET in the western US from our analysis. ET is typically sensitive to water availability in the western US, where is an arid region. Therefore, the overestimation of ET in GCM is attributed to the overestimated precipitation.

Soil moisture can influence the precipitation generation (*Koster et al.*, 2004; *Taylor et al.*, 2012a) through its impact on ET (*Gallego-Elvira et al.*, 2019). Specifically, local ET is an important source of water vapor for precipitation (*Burde and Zangvil*, 2001; *Li et al.*, 2018), especially for the arid regions where external water vapor advection is relatively low (*Dominguez and Kumar*, 2008). In turn, precipitation serves as the water supply for ET at water limit regions. The climate models used in this study trends overestimate ET Thus, an accurate estimation of ET may improve the performance of precipitation in the climate model.

Both the ET estimates from the MEP method and GCM outputs are projected to increase over the whole US continent in the future, but the magnitude of the

increase from the MEP method is lower. The smaller signal is likely because the MEP model includes soil moisture control on ET estimates in its formulation. Soil moisture was projected to be drier in the future, which represents increased water stress to suppress ET. Furthermore, projections from the MEP method are more robust, suggesting the uncertainty of the estimates is less, and agreement among the selected models are better. This can be attributed to the parsimony of the MEP theory, which requires only net radiation, specific humidity, temperature, and soil moisture as inputs. Other classical ET methods need inputs like wind speed, resistance parameter, water gradient, and surface roughness that are difficult to obtain and usually carry high uncertainty in model outputs. MEP method reduces such uncertainties without using that information, produces less uncertain ET estimates among climate models.

4.7 Conclusion

In this chapter, we validate a novel method of estimating ET based on the Maximum Entropy Production theory in the Amazon basin and apply it to regenerate ET with General Circular Model outputs at the US continent scale. While fewer inputs are required in the MEP method, it has higher performance in estimating ET in high biodiversity environment than the ET product used classical method by comparing to observations.

The MEP method is then applied to reproduce ET with outputs from GCMs, showing good skills of mimicking the averaged annual ET in the historical period as compared to the ET estimates using the water budget method. In terms of change of ET in the future, the multi-model ensemble of MEP-based projection shows higher robustness but relatively less increase than the projections of the GCM-based estimates. The uncertainty of MEP-based estimates is reduced compared to other methods due to the efficient use of available information about the ET process. It provides an

attractive alternative for monitoring ET at any scale without relying on inputs that hard to obtain or carry high uncertainty from reanalysis products, for example, water vapor gradient, wind speed, surface roughness, and stomatal resistance, etc. Further, an accurate ET estimate is useful for understanding the atmosphere and land surface interaction, which can improve the performance of precipitation estimations.

CHAPTER V

A Novel Framework of Urban Flood Simulation for Real-time Flood Forecasting Within Uncertainty Quantification Framework

5.1 Introduction

5.1.1 Flooding trends in the past

Major floods impact all regions of the world, including those with growing population and increased urbanization (Figure 5.1a). Despite significant advances in numerical weather prediction systems (*Benjamin et al.*, 2018) and improvements of models with forecasting capabilities, flooding in densely populated areas has remained the costliest natural hazard of all weather-related events in terms of fatalities (*Doocy et al.*, 2013). For example, in the U.S., while there has been a gradual negative trend since the late 1970s, the 30-year average of flood-related fatalities (1985-2018) remains high (Figure 5.1b). According to post-event analyses by the National Oceanic and Atmospheric Administration, about two-thirds of fatalities (1995-2017) are classified as driving and in water (Table 5.1). A common scenario is when a vehicle heads straight into waters, arguably, as a result of the lack of real-time information on the extent and flow conditions in the area of impact. Flooding also incurs one of the highest

material costs (*Winsemius et al.*, 2016) and more accurate, probabilistic mapping of areas prone to insufficient drainage, channel flooding, or coastal storm surges remains a high priority.

While the total number of annual floods worldwide exhibits a complex temporal trend, the proportion of global extreme floods is clearly on the rise (Figure 5.1c). Changes in flooding severity can be due to individual or combined changes in precipitation, land use, and cover, as well as practices of water management. These changes can co-occur, making it difficult to determine the relative importance of each factor as drivers of observed evolution in river flooding behavior (*Wyżga et al.*, 2016). Nonetheless, it is clear that the global hydroclimate is undergoing a transient state (*Fischer and Knutti*, 2016; *Kharin et al.*, 2013), and changes in flooding are consistent with the distribution of changes in precipitation extremes (*Mallakpour and Villarini*, 2015; *d'Orgeville et al.*, 2014; *Wasko and Sharma*, 2017).

5.1.2 Challenges of flood modeling in urban environments

Areas with high level of urbanization have particularly been hot-spots of flooding impacts because of the high density of population as well as a large fraction of impervious areas that can increase flooding severity. Recent estimates indicate that the number of people residing in the path of high-risk floods will double from one to two billion, within two generations *De Groeve et al.* (2015). Problems associated with flooding, therefore, have the potential to proliferate in the very near future rapidly. The emerging needs are both (i) to understand how urban environments affect the propagation of extreme floods, to inform mitigation measures, and (ii) to engineer comprehensive modeling capabilities relevant to decision making in the critical times of flooding. Both of these needs require spatially-explicit information on flood dynamics within the urban fabric that need to be commensurate with the typical scales at which humans operate, as they prepare or react to flooding. Such human action

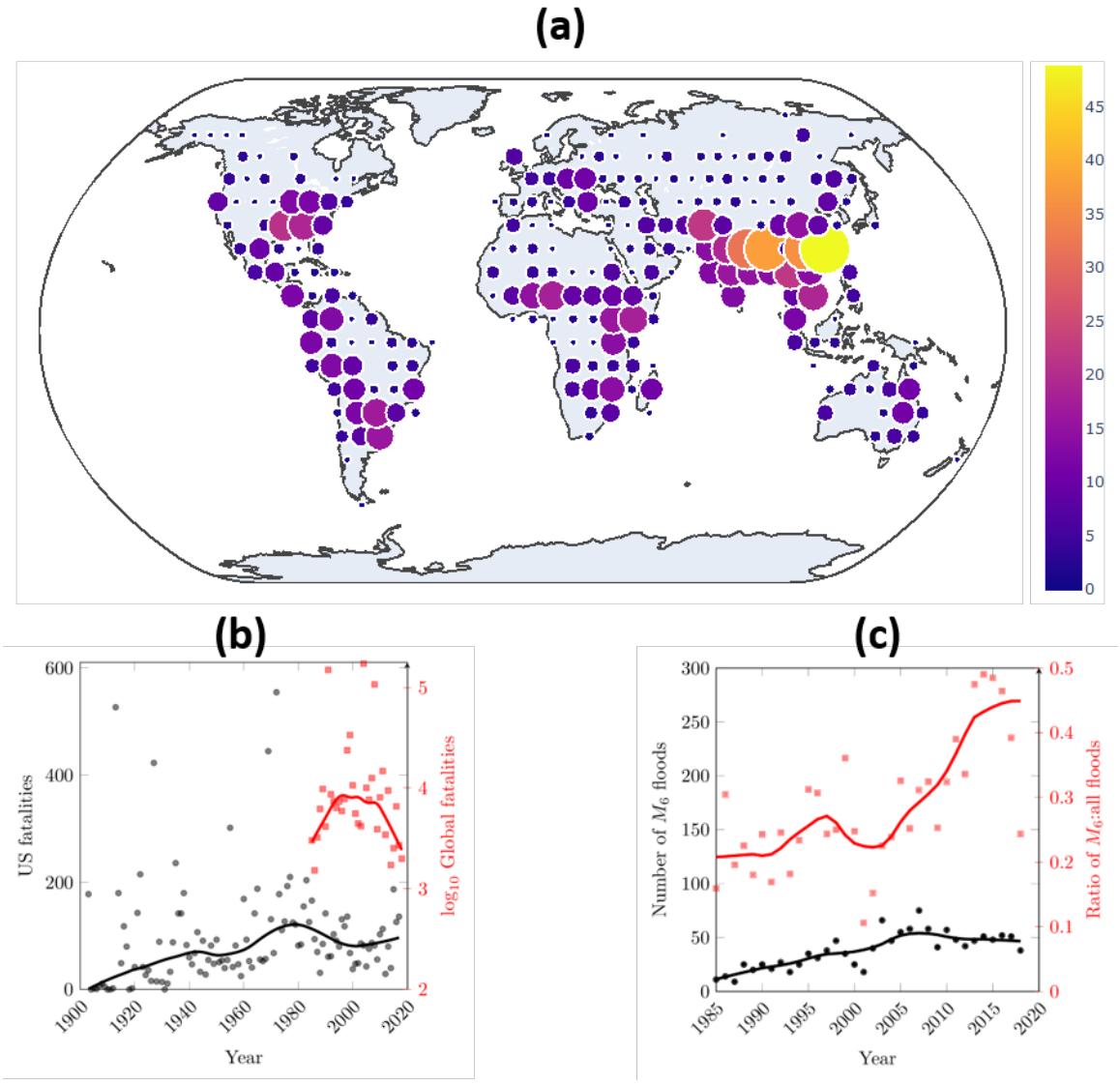

Figure 5.1: (a) Count of extreme floods for the period of 1985-October 2019. Symbol size is proportional to the count of floods occurring in a 7.5×7.5 degree grid. (b) Annual fatalities from flooding events: global (red squares) and the US only (black circles). All fitted lines are smoothed trends obtained through LOESS (*Cleveland and Devlin*, 1988) using weighted linear least squares. (c) The total number of annual extreme floods worldwide (black circles) and the ratio of extreme floods to all floods (red squares). Extreme floods are defined as events exceeding Flood Magnitude (FM) of 6, where FM is computed as the common logarithm of the product of flood duration, severity, and affected area (*G.R.Brakenridge*).

scales are arguably within the range that is characteristic of dimensions of a human dwelling, an evacuation route, or a vehicle (i.e., $\mathcal{O}(10^0) - \mathcal{O}(10^1)$ m). At these same scales, flooding information needs to contain a measure of quantified uncertainty, such

Table 5.1: Flooding fatalities partition of U.S. from 1995 to 20171

Year	Total	Vehicle	In Water	Flash Flood	River Flood
1995	80	39	35	60	20
1996	131	79	31	94	37
1997	118	46	45	86	32
1998	136	75	25	118	18
1999	68	26	29	60	8
2000	38	24	7	30	8
2001	48	24	12	35	13
2002	49	28	11	38	11
2003	86	39	21	67	18
2004	82	45	25	58	24
2005	43	18	19	28	15
2006	76	32	34	59	17
2007	87	50	21	70	17
2008	82	39	23	58	24
2009	56	33	10	33	23
2010	103	45	22	67	36
2011	113	68	26	69	44
2012	29	10	11	19	10
2013	82	37	17	60	22
2014	38	16	14	29	6
2015	187	112	24	129	45
2016	126	58	34	86	40
2017	136	40	52	N/A	N/A
Average	86.8	42.7	23.8	61.5	22.3
Percentage		49.25%	27.45%	70.87%	25.72%

***Note:** According to the NOAA report, most vehicular fatalities occur when people attempt to drive across flooded roadway. Of the deaths classified as being "in water", many of the victims were originally in vehicles when they stalled in high water, left the vehicle, were swept away by flood waters.

as confidence bounds, exceedance probability, or any other metric for which public and decision-makers might express preference (*Morss et al.*, 2008). Satisfying the requirements of the need calls for comprehensive modeling technologies and shifts in current paradigms of how flooding information is produced.

The history of modeling of runoff generation and overland flow has a rich history and dates back to the 1970s (*Buda*, 2013; *Fatichi et al.*, 2016). Recent developments have led to the integration of watershed hydrology and flow hydrodynamics in com-

prehensive, first principles-based approaches that can be used as 'flooding models' (*Kim et al.*, 2012a). A number of advances in understanding fluid motion and capabilities to represent various flooding characteristics (flow rate, water depth, dynamic pressure, inundation extent) in space and time have been achieved in the past two decades since the rapid advent of computational power. In urban areas, flood modeling has however remained difficult as human-made structures create complex mosaics of hydrologic conditions and flow impediments. Surface flow confluence and blockages due to buildings, bridge piers, flood control structures, and even non-uniformly distributed vegetation patches are typical situations that need to be holistically modeled. The representation of urban geometries to accurately resolve fluid object interactions requires refinement of the model mesh (*Schubert et al.*, 2008; *Schubert and Sanders*, 2012) that may result in excessively large number of computational elements (e.g., $\mathcal{O}(10^3)$ *per* km^2, see section 5.3.2). Despite typical simplifications of the governing mathematical formulations in flooding models derived from the Navier-Stokes equations describing the motion of the viscous fluid, the resultant partial differential equations require advanced modeling techniques, and their solution is still computationally expensive. Finer spatial resolutions might place additional constraints on the model time step to ensure numerical stability. As a result, the computational burden of flooding models may become prohibitively high even at moderate spatial scales of the application ($\mathcal{O}(10^3 - 10^4\ km^2)$). It is comparable to the typical of global circulation models used in climate projections ($\mathcal{O}(10^6 - 10^8)$ *cells*) that are run a moderate number of times offline in large computational centers. While the latter is not the ultimate example of computational complexity, they provide a useful benchmark because of the similarity of governing equations and scalability (see 5.4.3).

Even those flooding models that simplify the physics of fluid motion are challenged by the need to be executed in real-time, given forecast rainfall – even more so, when they need to assess the forecast uncertainty. Despite recent advancements

in the modeling of the underlying physics of inundation (*Bates*, 2012; *Bierkens et al.*, 2015; *Fatichi et al.*, 2016) and novel platforms and libraries offering efficient parallelized computations (*Artichowicz and Gsiorowski*, 2019; *Neal et al.*, 2010), real-time flood modeling at the required human action scale and with appropriate uncertainty quantification is still not feasible. Furthermore, conventional computational means available to a flood modeler – a workstation, a computational cluster, or even cloud computing infrastructure (e.g., (*Glenis et al.*, 2013)) – all remain to be modest and cannot address the above challenge. In real life, the lack of this computational capability transforms into the inability to make better-informed disaster management decisions, such as the selection of evacuation routes, determining areas to direct rescue resources, or risk analysis on buildings or critical infrastructure. In this chapter, we propose a novel, paradigm-changing framework for real-time flood modeling and uncertainty quantification that combines the rigor of classical physics of fluid motion and recent advances in applied mathematics.

5.1.3 A real-time flood forecasting framework

The framework contains four essential steps and is outlined in Figure 5.2. In step 1, the uncertain, crucial inputs of a flood-resolving, physical model of high fidelity are defined. Step 2 creates a set of surrogate models of reduced-order of complexity (e.g., polynomial functions, see section 5.2.1) from the response of the high-fidelity model forced by the inputs in step 1. As an optional step, one can perform inference (step 3a), given observed data that can be mapped onto the state-space of the flooding model. This can be done to learn about the effective properties of the modeled area and to provide more information (step 3b) about the probabilistic outcome of the uncertain inputs in Step 1. Once the set of surrogate models is created, real-time model inputs can be supplied (step 4) for fast computation. This enables real-time calculation (step 5) of relevant hydrologic features such as inundation, river discharge, and

water depth distribution that can be used for emergency planning or disaster response management. By decomposing the response of the high-fidelity physical model into a set of targeted quantities of interest (e.g., Figure 5.2), the framework offers progressive 'learning' of underlying physics with computationally cheap surrogate models. They add the benefit of probabilistic assessment of flooding variables in space and time and can be easily adjusted when new data arise. Importantly, the framework permits to move all intensive computations entirely into the period that is outside of the time window when real-time flooding information is needed.

One element of the proposed framework that contributes to a paradigm-changing solution for real-time flood forecasting, making it distinct from the traditional approaches, is the reallocation of computationally heavy physical simulations from the period of the flooding event to the time interval between major events (Figure 5.2, blue box). For example, in the area of study that would correspond to the non-hurricane season. These physical simulations of the flood-resolving model of high fidelity are used to 'pretrain' surrogate models representing specific quantities of interest in the flooded area such as water depth or velocity at a given location: a major road intersection, a freeway underpass, or floodplain. The reallocation of arduous calculations in time critically changes the requirements of computational resources: instead of needing large numbers of processing cycles during the flooding event when resources are likely to be limited, one can perform these computations prior to the event and store the simulations. Effectively, this replaces the challenge of computational burden with the problem of data storage in order to construct surrogates and ensure straightforward access to them something that can be successfully resolved with an ordinary desktop due to proliferation and low cost of data storage systems.

In this work, tRIBS-OFM, a physically rigorous hyper-resolution hydraulic and hydrologic model is used to simulate flood propagation and advance numerical representation and understanding of interactions between flooding and the urban en-

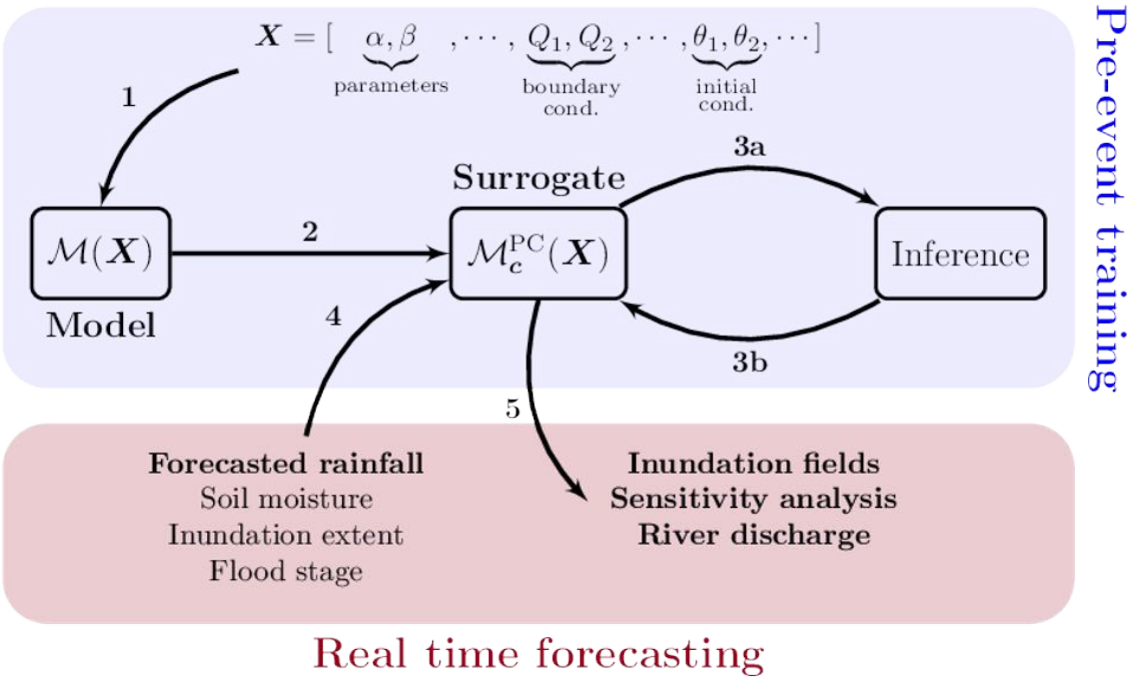

Figure 5.2: Workflow of a novel paradigm for real-time flood-forecasting with uncertainty quantification. A set of uncertain inputs X is required for a flood-resolving model $\mathcal{M}$ of high fidelity that can simulate spatially-varying structure of runoff generation and surface flow hydrodynamics in a domain of arbitrary complexity. A surrogate $\mathcal{M}_c^{PC}$ is a model of reduced order of complexity that is trained to represent any output quantity of interest (QoI) based on the response of the model $\mathcal{M}$ given its uncertain inputs X. During the construction of surrogates, observed data linkable to the outputs of $\mathcal{M}$ can be used to learn the likelihood of model uncertain parameters, resulting in parametric inference. All of the above computational effort needs to occur during a pre-flooding event period (blue box). This permits computational feasibility in real time (pink box), also resulting in fast propagation of uncertainties in inputs X (or parameters) through $\mathcal{M}_c^{PC}$ to get probabilistic estimates of QoIs. Pink box: bold face type denotes elements used in this study. Steps 1 through 5 outline connections among elements of the workflow.

vironments. Specifically, tRIBS is short for Triangulated Irregular Network based Real Time Integrated Basin Simulator, which is a comprehensive hydrologic model that is capable of reproducing runoff through several generation mechanisms (*Ivanov et al.*, 2004). OFM represents the overland flow model, solving 2D Saint-Venant equations with finite volume method and use Roes approximate Riemann solver for computing fluxes (*Bradford and Sanders*, 2002; *Begnudelli and Sanders*, 2006). *Kim*

et al. (2012a) further modified it to adapt for hydrological application (e.g., partially submerged mesh cell), and coupled it with tRIBS for routing runoff on the surface.

Any scalar output from tRIBS-OFM can be used as a quantity of interest (QoI), leading to the need to develop a surrogate model for each QoI. Many surrogate models can be trained to mimic various QoIs that are direct or derived outputs of the flood-resolving model both in space and time: flow rate, level, pressure, shear stress, etc. Once simulation results from a model of high fidelity are generated, surrogate models can be constructed and calculated nearly instantly even with a low-end computer at the time of the extreme weather event. Multiple frameworks fall into the class of surrogate models, e.g., Gaussian process (GP) models (*Rasmussen*, 2004), artificial neural networks (*Ripley*, 1996), support vector machines (*Abe*, 2010), and polynomial chaos expansions (PCEs) (*Xiu and Karniadakis*, 2002). These different classes of surrogates were often developed in parallel fields over the last two decades, with GP models and PCEs mainly being used in statistics and engineering (often computational fluid dynamics), whereas neural networks and support vector machines were more used in data-intensive applications. PCEs method is accepted here to train the surrogate model of tRIBS-OFM. It can capture the complex, non-linear behavior of the physical model through a learned polynomial expansion, and provide easily global sensitivity analysis (*Dwelle et al.*, 2019). The polynomial form of the PCEs guarantees the efficiency of surrogates, allowing for the uncertainty quantification to take place, even in real-time.

To test this framework, we consider a recent real-world example of the flooding that occurred during Hurricane Harvey in Houston, TX, in August 2017. This flooding event is reconstructed to emulate the efficacy of the proposed novel framework in the setting of a complex urban environment. In order to closely mimic an operational setting of the real-world scenario, we use quantitative precipitation estimates (QPEs) provided by NOAA water resources and forecasting professionals with an

18-hour lead time to construct uncertain rainfall forecasts. These forecasts serve as inputs to forecast probabilistic streamflow and flooding estimates from the models outputs. The efficacy of this approach is assessed by comparing derived inputs and forecast outputs to streamflow, water levels, and inundation pattern observed during this flooding event.

5.2 Methods

5.2.1 Polynomial Chaos Expansion

Polynomial Chaos Expansion (PCE) method is used in this chapter to construct surrogate model of the flooding model (i.e., tRIBS-OFM), which is able to simulate the behavior of the computationally complex model efficiently. Let's use $\mathcal{M}$ to denote the physical model with inputs $\boldsymbol{x}$, where $\boldsymbol{x} = x_1, x_2, \ldots, x_M$ and M is the total number of inputs. As a distributed model, the outputs of tRIBS-OFM are 2D in spatial and 1D in temporal scale. Nevertheless, if we select the output (e.g., $\hat{y}$) of a specified location at a specified time as our Quantity of Interest (QoI), $\hat{y}$ can be expressed as:

$$\hat{y} = \mathcal{M}(\boldsymbol{x}), \tag{5.1}$$

where $\hat{y}$ represents a scalar quantity. According to PCE theory, this physical model $\mathcal{M}$ can be expanded as sum of infinite multivariate orthogonal polynomials:

$$\mathcal{M}(\boldsymbol{x}) = \sum_{\boldsymbol{\alpha}} c_{\boldsymbol{\alpha}} \Psi_{\boldsymbol{\alpha}}, \tag{5.2}$$

$$\Psi_{\boldsymbol{\alpha}} = \prod_{i=1}^{M} \Psi_{i,\alpha_i(x_i)}, \tag{5.3}$$

where $\Psi_{i,\alpha_i(x_i)}$ is univariate polynomial whose form is associated with the potential distribution of uncertainty input variable x_i (see Table 5.2), and α_i is member of

multi-index $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_M)$. In practie, the Eq (5.2) can be approximated with a truncated PCE (*Xiu and Karniadakis*, 2002; *Lin and Karniadakis*, 2009) by only selecting terms with a total degree of polynomials smaller than a certain value p. This leads to a finite set $\mathcal{A}_p = \{\boldsymbol{\alpha} : \sum_{i=1}^{M} \alpha_i \leq p\}$ for the multi-index $\boldsymbol{\alpha}$ to take. For example,

$$\mathcal{M}(\boldsymbol{x}) \approx \mathcal{M}^{PC}(\boldsymbol{x}) = \sum_{\boldsymbol{\alpha} \in \mathcal{A}_p} c_{\boldsymbol{\alpha}} \Psi_{\boldsymbol{\alpha}}, \tag{5.4}$$

The above equation can also been written as:

$$\mathcal{M}(\boldsymbol{x}) \approx \mathcal{M}^{PC}(\boldsymbol{x}) = \sum_{j=0}^{P} c_j \Psi_j, \tag{5.5}$$

where j represents the order number of any possible multi-index $\boldsymbol{\alpha}$ in $\mathcal{A}_p$ in a predefined order (see details in Appendix B of *Dwelle* (2018)). There are $P+1$ polynomial basis and requires $P+1$ coefficients of c_j for constructing the PCE. This number is determined by both the input dimension M and the total degree for truncation p, and the relationship is derived in *Xiu and Karniadakis* (2002):

$$P + 1 = \frac{(M+p)!}{M!p!}, \tag{5.6}$$

This value P proliferates as the number of uncertain input variables increases. For example, in this work that we consider $M = 20$ uncertain input variables and truncate the PCE at $p = 3$, and it requires **1771** simulations of tRIBS-OFM to calculate the c_j in Eq (5.4) according to Eq (5.6). A typical urban watershed domain resolving at 10^0 - $10^1 m$ scale can have the number of computational cells at $\mathcal{O}(10^6)$ order. Solving tRIBS-OFM at such scale is very computationally expensive, and running **1771** simulations is extremely challenging. Therefore, we applied the Bayesian compressive sensing method (*Sargsyan et al.*, 2014) to construct PCE, which allows us to use much

fewer tRIBS-OFM simulations to find the coefficients c_j in Eq (5.5).

After computing of the PC coefficients c_j, the global sensitivity information can be extracted by computing the Sobol indices (*Sobol*, 2001). The main sensitivity index corresponding to input X_i can be estimated with the PC coefficients:

$$S_i = \frac{\sum_{j \in \Pi_i} c_j^2}{\sum_{j=0}^{P} c_j^2}, \tag{5.7}$$

where Π is the indices of basis terms that only involve input X_i, and P represents the total number of non-zero order basis terms. The Sobol indicesS_i can be interpreted as the fraction of variance in the output that is associated with the uncertain model input variable X_i only.

Table 5.2: Potential distributions for input variable and their orthogonal polynomials.

Distribution of x_i	Support	Polynomial
Uniform	[-1, 1]	Legendre
Gaussian	(-∞, ∞)	Hermite
Beta	[-1, 1]	Jacobi
Gamma	[0, ∞)	Laguerre

5.2.2 Karhunen-Loeve decomposition

Consider inundation depth to be a QoI simulated by tRIBS-OFM $\mathcal{M}(l;t;X)$, where l and t represent location and time. The uncertain inputs are denoted by $X \in \mathcal{R}^d$, where d is the dimensionality of the input space. To train the surrogate, we need to sample X S times and run tRIBS-OFM for each sample (e.g., X^i). If we have N_l cells in the mesh, N_t outputs in time, this gives $N_l \times N_t \times S$ forward model evaluations $\hat{y}_{ijk} = \mathcal{M}(l_j;t_k;X^i)$, which is training simulations with $i = 1,\ldots,S$, $j = 1,\ldots,N_l$, and $k = 1,\ldots,N_t$.

The approach described in section 5.2.1 aims to build a separate surrogate for each of the Quantity of Interests (QoIs). However, the high resolution of the domain

can result in a very large number of QoIs. Building and running a large number of surrogates may become computationally prohibitive. Instead, if the QoIs exhibit high spatial correlation, it is preferred first to reduce the dimensionality through Karhunen-Loeve (KL) decomposition:

$$\mathcal{M}(l;t_k;X) \approx \bar{\mathcal{M}}(l;t_k;X) + \sum_{m=1}^{M} \xi_m(X)\sqrt{\mu_m}\phi_m(l,t_k), \tag{5.8}$$

where $\bar{\mathcal{M}}(l;t_k;X)$ represents the mean field of training simulation at time t_k, $\xi_m(X)$ are KL coefficients, $\sqrt{\mu_m}\phi_m(l,t_k)$ are KL models scaled by the square-root of eigenvalues (e.g., $\sqrt{\mu_m}$), and $\phi_m(l,t_k)$ is the corresponding eigenvector, and m denotes the order of eigenvalues in descending order. Now, given training samples $\mathcal{M}(l_j;t_k;X^i)$ with k fixed at a specified time and $i = 1,\ldots,S$, $j = 1,\ldots,N_l$, the KL decomposition produces training samples for the coefficients $\xi_m(X^i)$, where $m = 1,\ldots M$. One needs to construct PCE surrogate for each of the M KL coefficients only. The determination of M depends on eigenvalues of the training samples. In general, $M \ll N_l$, and hence guarantee the computational efficiency.

From Eq (5.4), the surrogates of the KL coefficients can be expressed as:

$$\xi_m(X) = \sum_{j=0}^{P} b_{mp}\Psi_p, \tag{5.9}$$

Plugging into Eq (5.8) to retrieve original model outputs (e.g., inundation depth):

$$\begin{aligned}
\mathcal{M}(l;t_k;X) &\approx \bar{\mathcal{M}}(l;t_k;X) + \sum_{m=1}^{M}\sum_{j=0}^{P} b_{mp}\Psi_p\sqrt{(}\mu_m)\phi_m(l,t_k) \\
&= \bar{\mathcal{M}}(l;t_k;X) + \sum_{j=0}^{P}\left[\sum_{m=1}^{M} b_{mp}\sqrt{\mu_m}\phi_m(l,t_k)\right]\Psi_p \\
&= \sum_{p=1}^{P}\left[\bar{\mathcal{M}}(l;t_k;X) * \delta_{p,0} + \sum_{m=1}^{M} b_{mp}\sqrt{\mu_m}\phi_m(l,t_k)\right]\Psi_p,
\end{aligned} \tag{5.10}$$

which has a similar form as Eq (5.4) and the mean field is absorbed into the $0-th$ order coefficient via the Kronecker delta $\delta_{p,0}$,

$$\delta_{i,j} = \begin{cases} 0 & if \quad i \neq j \\ 1 & if \quad i = j \end{cases} \tag{5.11}$$

5.2.3 Error index

Relative error norms are used to measure the differences between tRIBS-OFM and surrogate model:

$$\begin{aligned} \epsilon_q &= \frac{\left\| \mathcal{M}(X) - \mathcal{M}^{PC}(X) \right\|_2}{\left\| \mathcal{M}(X) \right\|_2} \\ &= \frac{\sqrt{\sum_{i=1}^{S} (|\mathcal{M}(X^i) - \mathcal{M}^{PC}(X^i)|)^2}}{\sqrt{\sum_{i=1}^{S} \mathcal{M}((X^i)^2)}}, \end{aligned} \tag{5.12}$$

where q is the index for a QoI, and S is the total number of training simulations. This index evaluates the averaged error for the PCE surrogate over all training data. This index can also be used to measure the error of the surrogate model in validation simulations.

5.3 Data

5.3.1 Study domain

Houston 2017 August flooding event is selected to demonstrate the uncertainty quantification framework for real-time flood-forecasting problem. A watershed monitored by United States Geological Survey (USGS gauge#08074540) is identified in the city as the case study domain. Subplot (a) in Figure 5.5.1 shows the landuse of Houston city and for the location of the selected watershed, inside which

illustrates a high fraction of developed areas. The land use information is downloaded from National Land Cover Database 2016 (https://www.mrlc.gov/data?f%5B0%5D=category%3Aland%20cover). Subplot (b) illustrates the inundation depth map at 3m resolution developed by the Federal Emergency Management Agency (FEMA). This map has been used to analyze the flooding extent and determine damage levels on specific structures (see details at https://doi.org/10.4211/hs.165e2c3e335d40949dbf501c97827837), but we need to note this product was only available after the event, and does not have an associated time of occurrence.

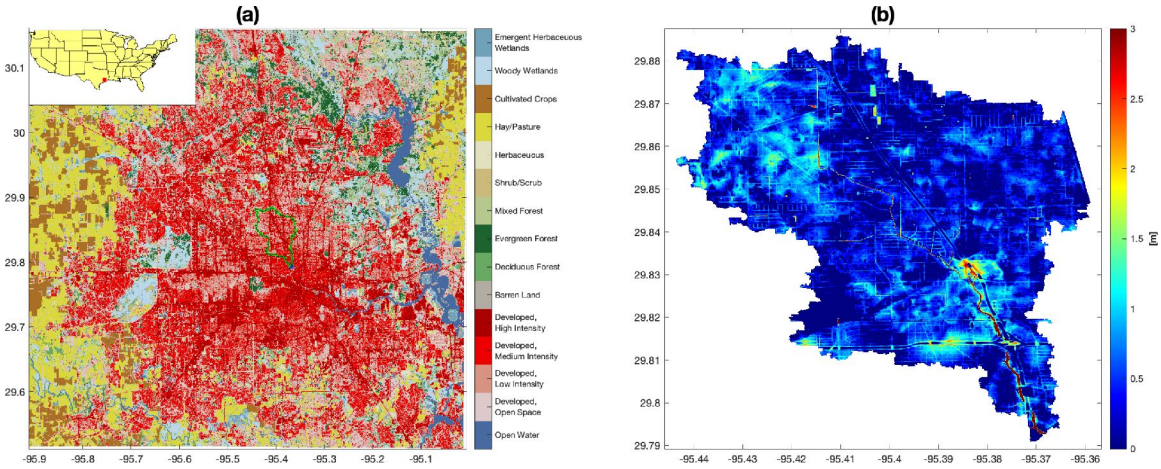

Figure 5.3: Study watershed in the greater Houston area. (a) Land use of Houston city and the watershed is identified with a yellow line; (b) is FEMA inundation depth for the watershed.

5.3.2 Mesh generation

The mesh of the study watershed is generated based on the 3m Digital Elevation Model (DEM) from USGS (available at https://viewer.nationalmap.gov/basic/), building footprints, stream networks, and transportation paths. The procedure is illustrated in Figure 5.4. Specifically, ArcGIS tool "Raster2TIN" is used to generate Triangulated Irregular Network (TIN) from the 3m resolution Digital Elevation Model (DEM) from USGS. TIN is the representation of the terrain surface with triangular facets that can capture the terrain information with fewer nodes

as compred to the original DEM. To understand the effects of buildings to flood wave propagation, we next merge the building footprints (download from `https://koordinates.com/layer/12890-houston-texas-building-footprints/`) into the TIN nodes. The building footprints are simplified by ArcGIS tool "Simplify Polygon" to some extent to make the buildings' shape more regular. Figure 5.5 illustrates a small region of the simplified buildings. There are two major reasons for the simplification: 1) avoid the occurrence of tiny triangles (area $< 0.1\ m^2$), as the smallest area of computational cell restricts the time step in finite volume method for stability; and 2) reduce the total number of cells and sequentially reduce the computational burdens of the simulation.

Usually, the stream network can be extracted from the DEM based on the surface flow accumulation. However, in our study case, the stream becomes discontinuous where a highway/bridge crosses the rivers. To fix these discontinuities, we add additional stream nodes discretized from continued streamline to the TIN nodes and simplified building footprints. Additionally, bridges and transportation paths nodes are added to keep the connectivity of the urban environments. Finally, all the nodes are input into a Delaunay Triangulator (available at `https://www.cs.cmu.edu/~quake/triangle.html`) to generate a constrained Delaunay triangular mesh (*Shewchuk*, 1996).

In summary, the watershed area is 46.9 km^2, and results in 136,423 mesh nodes and 271,215 triangle cells. There are about 20,000 buildings included in the mesh, which makes up 30% of the total watershed area. All of the land cover is developed areas: open space (13.2%), low intensity (31.6%), medium intensity (34.6%), and high intensity (21.1%).

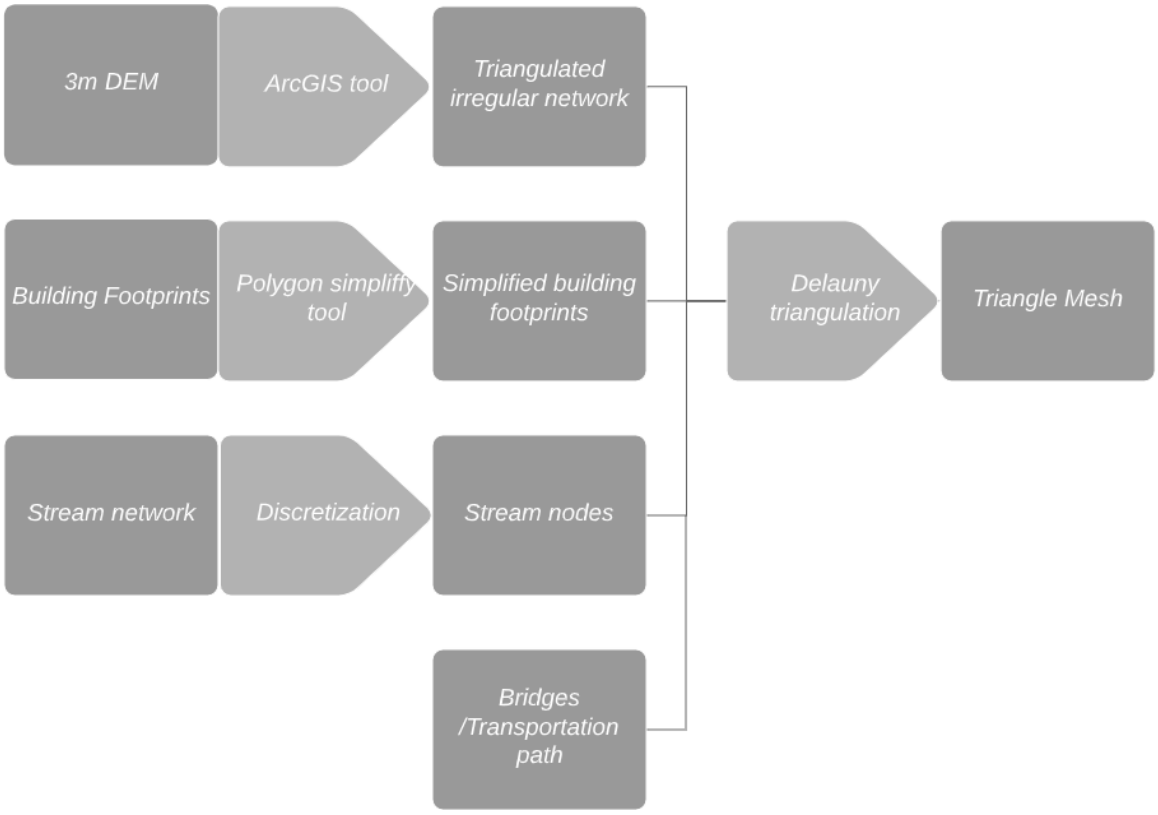

Figure 5.4: Mesh generation process.

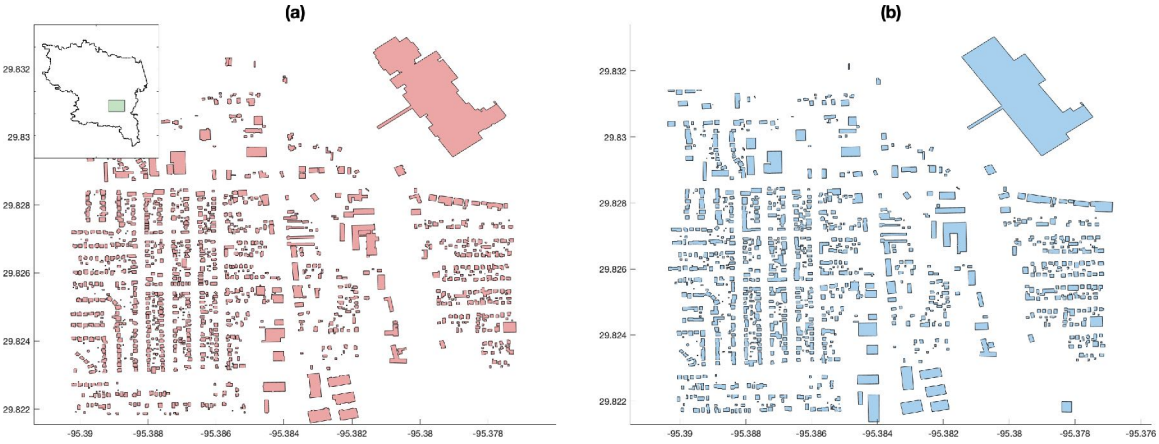

Figure 5.5: An example of building simplification. Subplot **(a)** illustrates the original building footprints for the green area in the top left insert plot, and **(b)** represents the simplified building shapes.

5.3.3 Precipitation

Another challenge in urban flooding forecasting is the difficulty of measuring or forecasting rainfall at a high resolutions and accuracy. Figure 5.6 shows the accumulated rainfall at the study watershed for the simulated period (Aug 27 00:00 Aug 27 5:00 AM, 2017) from 4 sources: 1) radar product ($500m \times 500m$); 2) Precipitation Measurement Mission (PMM ($0.1° \times 0.1°$), download from `https://pmm.nasa.gov/data-access/downloads/gpm`); 3) High-Resolution Rapid Refresh forecast product (HRRR ($3000m \times 3000m$), `https://rapidrefresh.noaa.gov/hrrr/`); 4) results from Regional Spectral Model ($12km \times 12km$, assimilating 8-km CMORPH (interpolated to a 12-km resolution grid), with initial and boundary conditions from NCEP-DOE R2; 5) gauge station from Weather Underground (Station ID: KTXHOUST1941, available at `www.wunderground.com/weather/us/tx/houston/KTXHOUST1941`). This station is not inside the watershed, but it is the nearest available precipitation gauge. Figure 5.7 illustrates the spatial rainfall distribution in the watershed from the four spatial products at the same time. Both figures highlight the significant uncertainty of the rainfall data at local scale, which should be considered when they are used to forecast flood events.

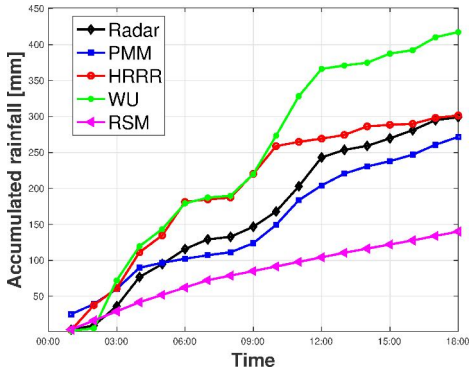

Figure 5.6: Accumulated rainfall in the watershed from Radar product, Precipitation Measurement Mission, High-Resolution Rapid Refresh, and Weather Underground. The X-axis represents the time of Apr 27th in GMT.

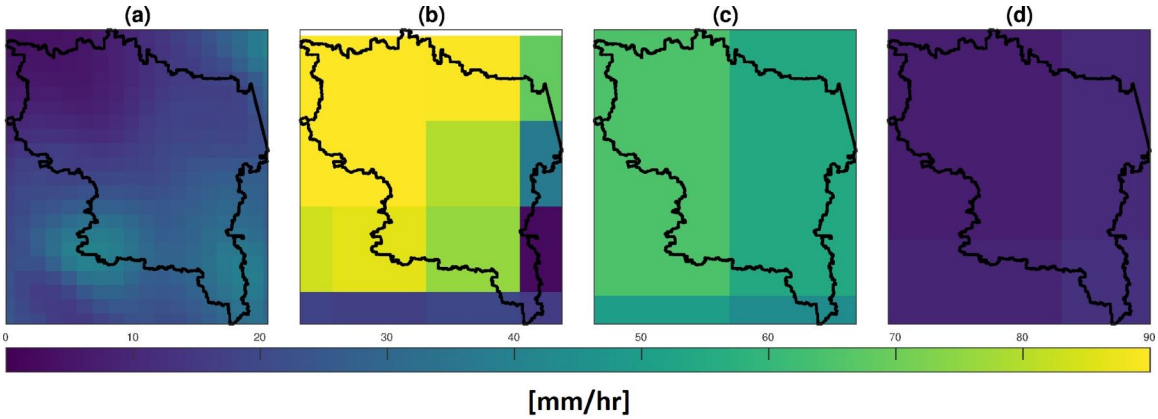

Figure 5.7: Snapshot of spatial distribution of rainfall intensity at Apr 27th 5:00 AM from (a) Radar product, (b) Precipitation Measurement Mission, (c) High-Resolution Rapid Refresh, and (d) Regional Spectral Model. All the three subplots are plotted at the same color scale, which is shown in the left panel and the unit is [mm/hour].

5.4 Overland Flow Model (OFM)

5.4.1 Governing equations

The overland flow model solves the 2D Saint-Venant equations (e.g., shallow water equation *Begnudelli and Sanders*, 2006), which derived from depth-integrating the Navier-Stokes equations describing the motion of viscous fluid. The governing equations consist of a continuity equation and two momentum equations for two perpendicular horizontal directions:

$$\frac{\partial \mathbf{U}}{\partial t} + \frac{\partial \mathbf{E}}{\partial x} + \frac{\partial \mathbf{G}}{\partial y} = \mathbf{S}, \tag{5.13}$$

where $\mathbf{U}$ is vector of flow variables, $\mathbf{E}$ and $\mathbf{G}$ are the flux terms in x and y direction respectively, and $\mathbf{S}$ is the source vector (*Kim et al.*, 2012a). Specifically,

$$\mathbf{U} = \begin{bmatrix} h \\ uh \\ vh \end{bmatrix},$$

$$\mathbf{E} = \begin{bmatrix} uh \\ u^2h + \frac{1}{2}gh^2 \\ uvh \end{bmatrix}, \mathbf{G} = \begin{bmatrix} vh \\ uvh \\ v^2h + \frac{1}{2}gh^2 \end{bmatrix}, \tag{5.14}$$

$$\mathbf{S} = \begin{bmatrix} i \\ -gh\frac{\partial z_b}{\partial x} - C_D u\sqrt{u^2+v^2} \\ -gh\frac{\partial z_b}{\partial y} - C_D u\sqrt{u^2+v^2} \end{bmatrix},$$

where h represents flow depth, u, v are the flow velocity in x and y directions, g is the gravity acceleration, constant z_b is bed elevation, and C_D is surface drag coefficient. A detailed numerical scheme of the numerical solution with the finite-volume method can be found in *Begnudelli and Sanders* (2006).

Solving a full set of the 2D Saint-Venant equations is valid for any kinds of channel conditions, but remains numerically challenging. Further, the time step can be constrained by the smallest cell area in the mesh in the finite-volume method for numerical stability, which introduces another challenge in the forecast with the overland model. For example, a typical time step is about $\mathcal{O}(10^{-1})$ for a high resolution application (e.g., cell area is $\mathcal{O}(10^{1})$).

5.4.2 Code verification

The parallelization of OFM is accomplished using Portable, Extensible Toolkit for Scientific Computation (PETSc; *Balay et al.*, 2019) and the PDEs are solved with the forward Euler scheme (*Abhyankar et al.*, 2018). Code verification is done with the Method of Manufactured Solutions (MMS *Roache*, 2001). Specifically, a combination

of exponential and sinusoidal functions is used for an analytic manufactured solution:

$$
\begin{aligned}
h(x,y,t) &= h_0(1+\sin(\frac{\pi x}{L_x})\sin(\frac{\pi y}{L_y}))\exp(t/t_0),\\
u(x,y,t) &= u_0(1+\sin(\frac{\pi x}{L_x})\sin(\frac{\pi y}{L_y}))\exp(t/t_0),\\
v(x,y,t) &= v_0(1+\sin(\frac{\pi x}{L_x})\sin(\frac{\pi y}{L_y}))\exp(t/t_0),
\end{aligned}
\tag{5.15}
$$

where h_0, u_0, and v_0 are the initial values for flow variables, L_x, L_y and t_0 are constants. Then, the derivatives that appear in the governing equation 5.13 can be evaluated exactly as presented in Appendix C. A regular grid is used to implement the code verification (Figure 5.8).

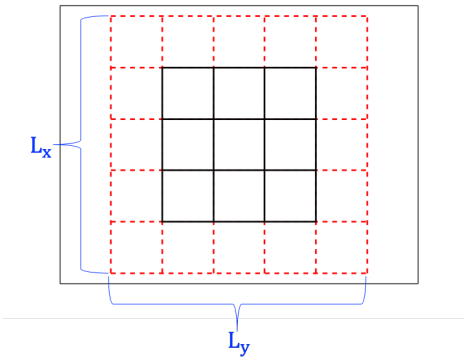

Figure 5.8: Regular grid for code verification. The black solid lines are the inner cells and the red dashed lines represent the boundary cells. $L_x = L_y = 5m$

The time step dt is fixed at constant value and allow the spatial resolution to vary as within a range $dx = dy \in [0.05, 0.1, 0.25, 0.5, 1]$ meter to test the model error associated with spatial discretization. The time step $dt = 0.01s$ is selected to make sure the error is dominated by the spatial discretization. Other parameters are setup as: $h_0 = 0.005$, $u_0 = 0.025$, $v_0 = 0.025$, $t_0 = 20$. The initial condition is specified by inserting $t = 0$ in Eq. (5.15), and the Dirichlet condition (e.g., constant values are specified at the boundary) is chosen as the boundary condition. The code is run for $t_{end} = 5s$, thus, the exact solution is $h(x, y, t = 5)$, $u(x, y, t = 5)$, and $v(x, y, t = 5)$.

The relative errors L_1, L_2, and L_{max} are used to measure the error of solution from running the code:

$$L_1 = \|h_{exact} - h_{ref}\|_1 = \frac{1}{N}\sum_{n=1}^{N}|h_{exact,n} - h_{ref,n}|,$$
$$L_2 = \|h_{exact} - h_{ref}\|_2 = (\frac{1}{N}\sum_{n=1}^{N}(|h_{exact,n} - h_{ref,n}|)^2)^{1/2}, \tag{5.16}$$
$$L_{max} = \|h_{exact} - h_{ref}\|_\infty = max|h_{exact,n} - h_{ref,n}|,$$

were h_{exact} is the exact solution from Eq (5.15), and h_{ref} represents the solution from the code.

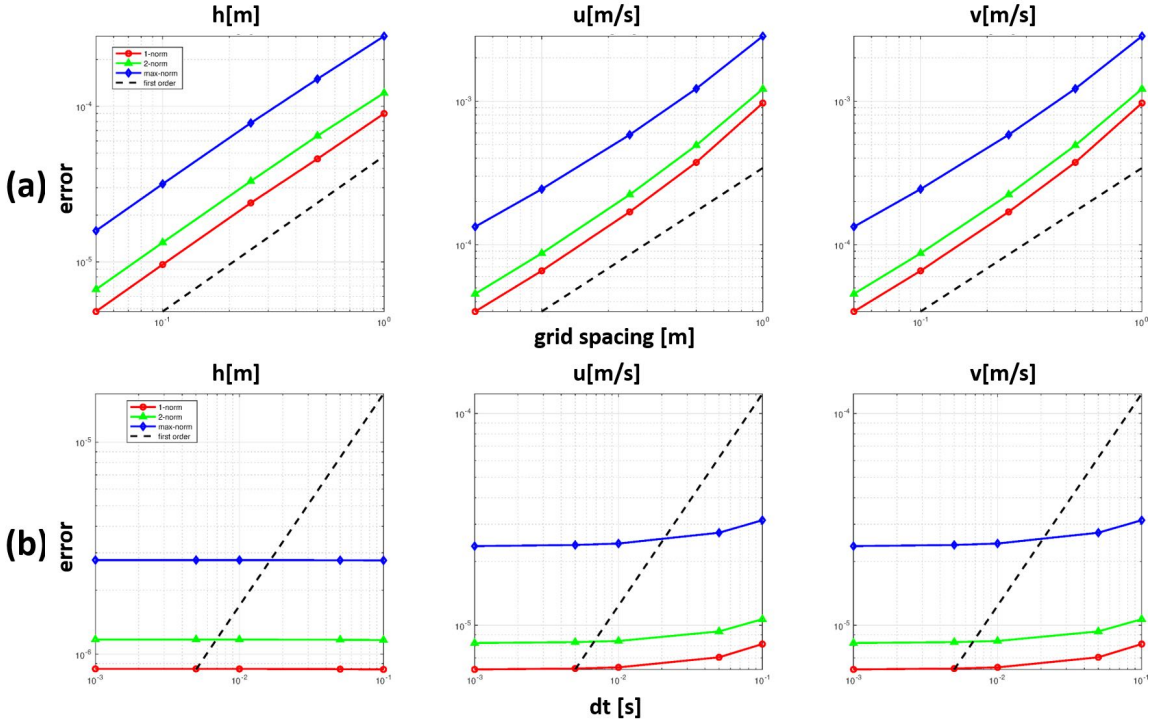

Figure 5.9: Errors of the code for the three variables, such as water depth (left panel), x-direction velocity (middle panel), and y-direction velocity (right panel) from different (a) spatial discretization and (b) temporal discretization. The red line, green line, and blue line represent the error in L_1 norm, L_2 norm, and L_{max} norm, respectively. The blue dashed line is the first order trend of the error decrease by reducing the grid spacing and time step.

As shown in Figure 5.9a, the error is proportional to the grid spacing. This

behavior is consistent with the finite volume method we implement in the code for solving the 2d Saint-Venant equations (*Kim et al.*, 2012a). However, the error cannot be reduced by fixing $dx = dy = 0.1[m]$ and changing dt from $0.1[sec]$ to $0.001[sec]$. The possible reason is that the time step is constrained by the cell area to guarantee numerical stability (e.g., the Courant-Friedrichs-Lewy condition). To guarantee a stable solution, the selection of time step dt always results in much smaller error as compared to the spatial discretization. Thus, the error from spatial discretization dominates in the total error, and the decrease of the error is negligible with a smaller time step.

5.4.3 Computational performance and scalability

Global Circulation Models (GCMs) or Earth System Models (ESMs) are used in climate projections with mesh representing the Earth system with $\mathcal{O}(10^6 - 10^8)$ cells. GCMs are run a moderate number of times, typically less than a few dozens of times, offline in large computational centers. While GCMs are not the ultimate examples of computational complexity, they provide a useful benchmark because of the similarity of governing equations and scalability as compared to flooding models of high fidelity.

There are numerous ways to assess computational performance of a numerical model. To enable a cross-comparison with models of similar as well as reduced-order of complexity, we introduce Core Seconds per Simulated Day - Normalized (CSSD-N):

$$\eta = \frac{\tau_{SD}}{N}, \tag{5.17}$$

where τ_{SD} represents the total CPU seconds required for a 24-hour simulation period and N is the total number of computational cells used by the model. As the formulation suggests, η is a performance metric normalizing simulation time per unit cell, which attempts to eliminate the effect of dimensionality of the modeled problem.

There are $271,215$ cells in the case study domain, and ~ 52 hours are used to simulate a 24-hour flood event with the model run in a serial mode (Intel(R) Xeon(R) Gold 6140 CPU /@ 2.30GHz). This yields a value of the CSSD-N for the flood-resolving high-fidelity model (FHFM):

$$\eta_{FHFM} = \frac{52 \times 3600}{271,215} = 0.69 \quad [sec], \tag{5.18}$$

or, more generally, the typical order of η_{FHFM} is expected to be in the range $\mathcal{O}(10^{-}1)$ - $\mathcal{O}(10^{0})$. Although inundation simulation is typically carried out at fairly small scales of a local watershed, just as illustrated in the case study, it might have a larger computational burden as compared to an application of a Global Circulation Model (GCM). This is because urban floods need to be simulated at a high spatial resolution (few meters) and, correspondingly, small time steps are required for numerical stability. For example, to evaluate computational performance of GCMs in the Coupled Model Intercomparison Project 6 (CMIP6) *Balaji et al.* (2017) used Core Hours per Simulated Year (CHSY) defined as the product of run time for one simulated year and the number of cores allocated to the simulation. Typical grid dimension of a GCM varies from $\mathcal{O}(10^{6})$ to $\mathcal{O}(10^{8})$ and it takes from $\mathcal{O}(10^{2})$ to $\mathcal{O}(10^{5})$ core hours to finish a simulation of 1 year period. To provide a comparison with the simulation performance for an urban flooding problem, we convert CHSY metric to CSSD-N using the following transformation:

$$\eta_{GCM} = \frac{\mathbf{CHSY}}{N_{GCM}} \frac{1}{365 \times 3600}, \tag{5.19}$$

where N_{GCM} denotes the number of computational cells used in a GCM. Values of CHSY and N_{GCM} reported in *Balaji et al.* (2017) lead to GCM varying from $\mathcal{O}(10^{-4})$ to $\mathcal{O}(10^{-2})$ which are 2-3 orders of magnitude lower than η_{FHFM} estimated above.

One needs to note that GCM simulations are run off-line (no real-time applica-

tions are intended), and this is typically done at high-performance computing (HPC) centers using hundreds to thousands cores with a parallelized version. While the computational performance of a high-fidelity flooding model is quite poorer than that of a typical GCM, it is unlikely to expect that the problem can be resolved using a similar 'brute-force' approach, i.e., by allocating many CPUs for flood simulation in real time. Model parallelization can shorten the execution time, but there is evidence that simulation will still not be fast enough for real-time forecasting as the speedup gain is limited (see Figure 5.10; the scalability of the model used in this study is similar to others reported in *Artichowicz and Gsiorowski* (2019) and *Neal et al.* (2010)). From a practical point of view, modeling every urban center subjected to flooding during an extreme event using a large number of cores would create a deficit of computational resources. Furthermore, no formal uncertainty quantification would be still possible as the latter requires many thousands of forward model simulations.

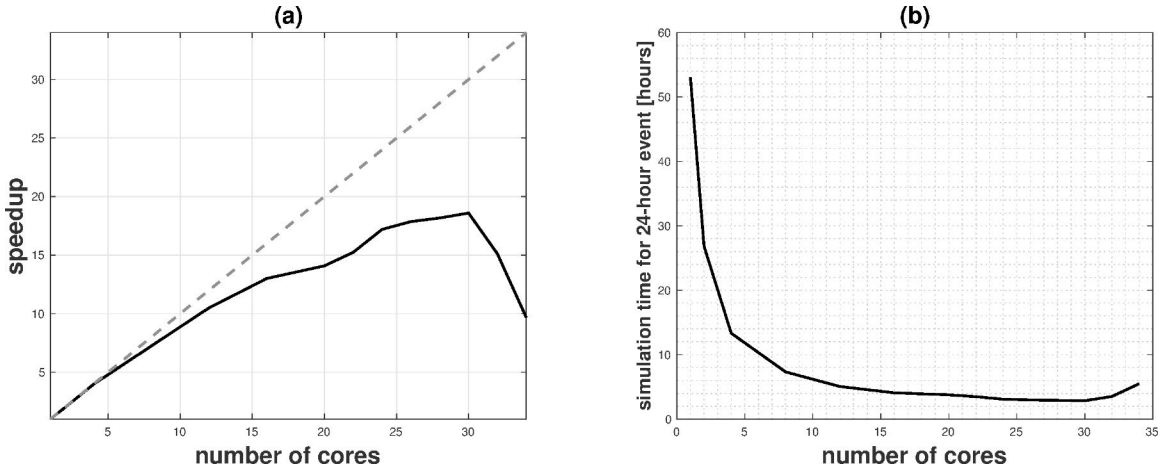

Figure 5.10: Computational performance of parallel OFM model with subplot (a). speedup and subplot (b). simulation time for simulating 24-hour event.

5.5 Results

5.5.1 Validation of tRIBS-OFM

The flooding model needs to be validated before implement the uncertainty quantification framework. Figure 5.11(a) shows a comparison of the discharge at the outlet between tRIBS-OFM forced with the Radar precipitation product (see Figure 5.7) and the USGS observed streamflow. The model simulates the rising limb of the hydrograph relatively well but not able to capture the recession curve. Specifically, the model simulated discharge declines faster than the observation after the peak. This may be caused by the lower estimate of rainfall after hour 12 in the simulation period, which is not sufficiently high to maintain the simulated streamflow at magnitudes of the observation. Thus, it is crucial to address the rainfall data uncertainty in flood modeling.

Besides the need to account for the uncertainty in model inputs, one needs to recognize uncertainty in observational data. Streamflow also carries uncertainty, making it a challenge to validate the model results (*Kiang et al.*, 2018). For example, the calibration limit for the monitoring station is 2.83 m^3/s, which is 2 orders of magnitude lower than the reported peak flow discharge during the event. Further, the Harvey event was extremely rare, and the maximum 3-days precipitation was over 1000 mm (*van Oldenborgh et al.*, 2017; *Emanuel*, 2017), This unusually high precipitation may have filled up adjacent watersheds/depressions/ponds, and they would spill the water to the study domain. Such additional water sources cannot be tracked since we only focus on the case study watershed. The larger domain should be simulated to study this possibility, which is beyond the scope of this study.

Figure 5.11(b) illustrates the inundation depth of the whole watershed at the time of peak discharge (e.g., hour 13). We zoom in three regions to present details of inundation caused by different mechanisms. A highway passes through region

1 horizontally, which is an area of particular interest for forecast during the event because people need to evacuate through it. Region 2 is a commercial district where the river passes through and represents the example of fluvial flooding. Region 3 is a residential area where inundation caused by heavy precipitation (e.g, pluvial flooding). Unlike the discharge is monitored by USGS gauge, the inundation depth is not always available. The only inundation depth data for our selected period and our study domain, to our best knowledge, is the FEMA map shown in Figure (b). However, this product is a post-event accumulated inundation estimate that was generated by interpolating DEM, high watermarks, and other model results. It has been used to evaluate the damage of flooding to buildings, but cannot be used to validate the simulation.

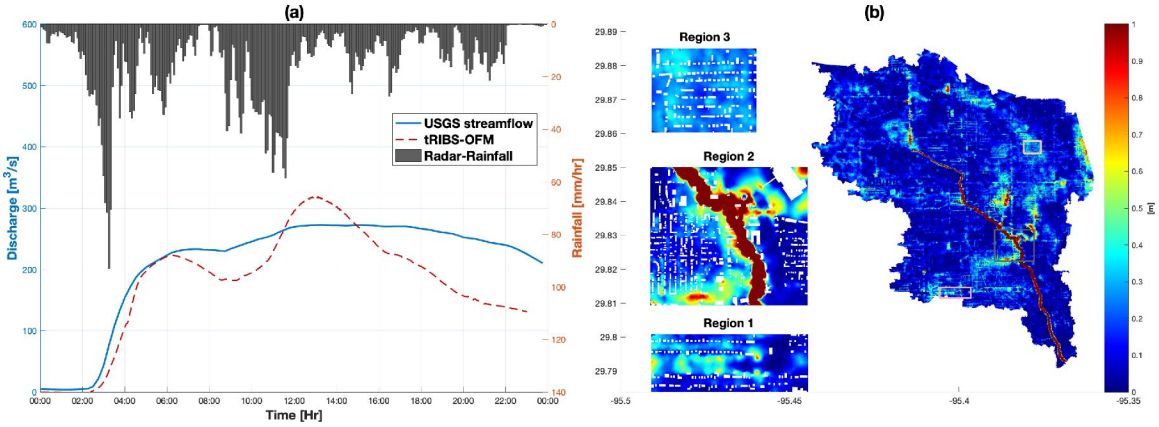

Figure 5.11: Validation of tRIBS-OFM simulation results. For the model setup, zero depth initial condition, saturated initial soil moisture, and spatial uniform roughness are assumed. Subplot (a) shows the comparison of the discharge at the outlet between tRIBS-OFM forced by the radar precipitation (red dashed line) and the observed streamflow from USGS gauge (solid blue line). Basin-averaged precipitation from the used Radar product is plotted on the top of subplot; (b) illustrates the snapshot of inundation depth on August 27, 13:00 from the results of tRIBS-OFM forced by the radar precipitation product. Three regions are blown up to show details corresponding to the three boxes in the watershed, such as region 1: yellow box, region2: black box, and region 3: pink box. The white areas represent the building footprints.

5.5.2 Performance of surrogate model

The surrogates were trained using series of uncorrelated pulses to capture. Rainfall rate within each hour is assume to follow the uniform distribution within the limits of the distribution.Specifically, $Pr = [X_1, \ldots, X_{18}]$ and

$$X_i \sim U(0, r_{max}), \quad i = 1, \ldots, 18 \tag{5.20}$$

where r_{max} represents the maximum hourly precipitation rate that is dependent on the climate of the study area. The process defined by Eq. (5.20) includes all the possible 18-hour rainfall series. For the study domain, $r_{max} = 50\, mm/hr$ is sufficiently high enough to capture the rainfall variability. Figure 5.12 shows examples of 100 realizations from Eq (5.20). In practice, we feed 1000 random 18-hour rainfall series to tRIBS-OFM to simulate the training sets for reduced-order modeling. Then the 1000 training simulations are used to construct surrogates for QoIs.

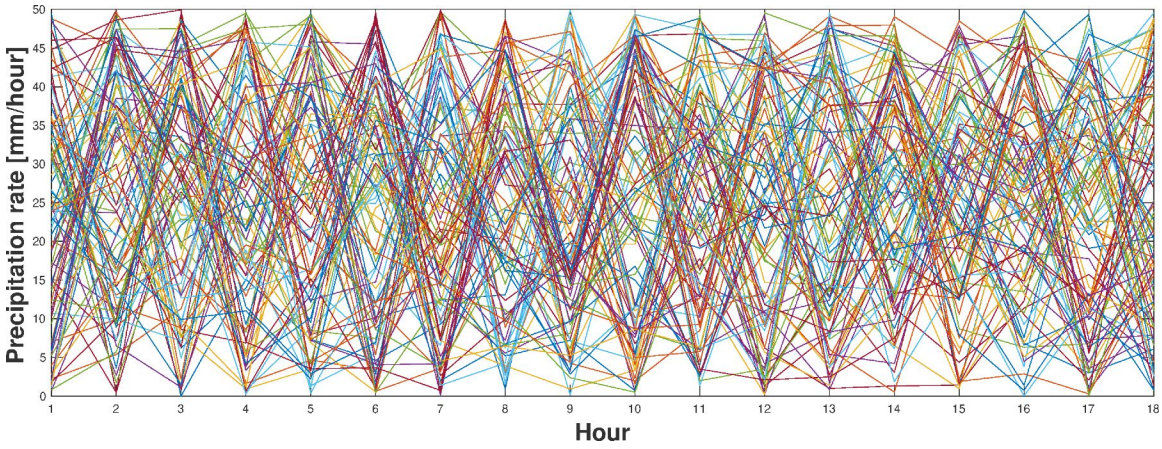

Figure 5.12: An example of realizations of rainfall represented as series of uncorrelated pulses from Eq. (5.20) (100 realzations).

In this chapter, we select the half-hourly discharge at the outlet, hourly inundation depths of all the cells in the three regions (zoomed in) as visualization in Figure 5.11 for QoIs to construct PCE based surrogates. Specifically, we truncate PCE at order $p = 3$, and train the polynomial coefficients with Bayesian compressive sens-

ing method (*Sargsyan et al.*, 2014). Each surrogate model needs to be trained with 1000 training simulations individually. The study has 36 surrogates for discharge and $(1,292+4,476+1,289)\times 18 = 127,026$ surrogates for inundation depth. Specifically, $1,292$, $4,476$, and $1,289$ are the numbers of cells in Region 1, Region2, and Region 3, respectively and 18 hours represent the duration of the simulation period for each location from tRIB-OFM. The surrogate's performance is validated with 150 validation simulation, which are forced with additional random rainfall series generated using Eq (5.20) (see Figure 5.13). The validations for inundation depths at other hours have good performance as well as Figure 5.13(b), (c), (d), but are not shown here. In brief, the average error for all the surrogates are **0.015**, **0.072**, **0.009** for Region 1, Region2, and Region 3, respectively.

To understand the reason for why surrogates have a relatively higher error in Region 2, we present training inundation depths of 4 QoIs from lower error to higher error in terms of surrogate performance (Figure 5.14). Specifically, subplot (a) represents an example of surrogate with good performance, and the relationship between the depth and antecedent rainfall is almost linear. Such relationships in Figure 5.14(b), (c), (d) exhibit thresholds: when the accumulated antecedent rainfall reaches some value, there is an abrupt change in the relationship slope. This phenomenon of nonlinearity can be caused by the complexity of local topography and building connectivity. For example, Figure 5.15 illustrates the the elevation contours of topography near the location of QoI of Figure 5.14(c). The small depression on the west side of this region introduces additional contributing areas by spilling water to this QoI location when it is filled up by antecedent rainfall.

In summary, the examples of inferior surrogate performance can occur anywhere in the domain, depending on topography and building effects. Table 5.3 shows that Region 2 has more QoIs behaving as those shown in Figure 5.14, which results in larger overall surrogates' error.

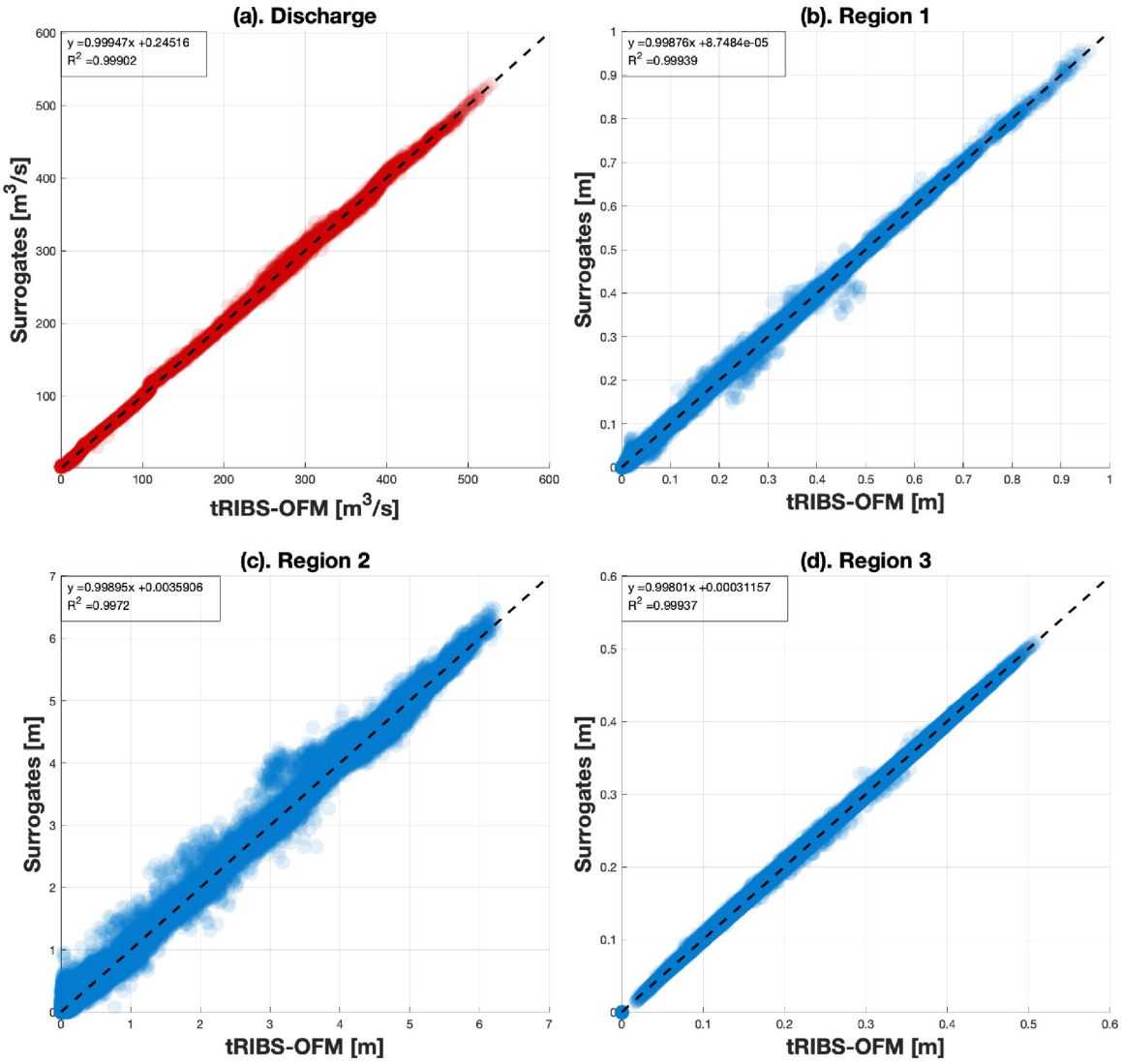

Figure 5.13: Validation of surrogate model with 150 validating simulations for (a) 36 half-hourly discharges; and inundation depth of all cells at hour 14 in (b) Region 1, (c) Region 2, (d) Region 3. The X-axis represents the tRIBS-OFM simulation results, and Y-axis denotes the values from the corresponding surrogate model. The black dashed line is the reference line of $y = x$. The fitted linear regression and R^2 are inserted in the left corner of each subplot.

5.5.3 Reduction of QoI dimensionality

In many instances with flood modeling, there can be many QoIs to be investigated. This could be flooding depths at locations in the study domain, at certain times in the simulation. For example, we select the hourly inundation depth of $7,057$ cells as our QoIs, which leads to $7.057 \times 18 = 127,026$ surrogates to train in total. Training

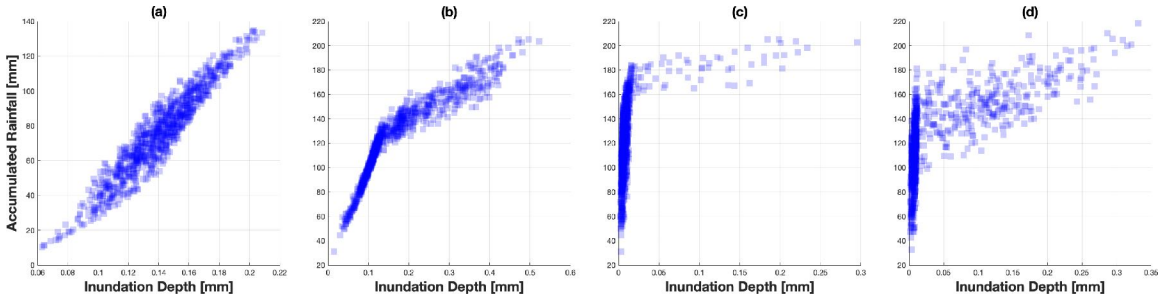

Figure 5.14: Relationship between training simulation inundation depths of a single cell and accumulated antecedent rainfall for surrogate error equal to (a). 0.01, (b). 0.1, (c). 0.2, and (d) 0.3.

Table 5.3: Distribution of PCE surrogates errors over different ranges.

	$\epsilon_q \leq 0.01$	$0.01 < \epsilon_q \leq 0.1$	$0.1 < \epsilon_q \leq 0.2$	$0.2 < \epsilon_q \leq 0.3$	$\epsilon_q > 0.3$
Region 1	53.36%	45.88%	0.62%	0.13%	0.026%
Region 2	13.73%	60.10%	20.67%	4.72%	0.77%
Region 3	63.37%	36.63%	0.009%	0%	0%

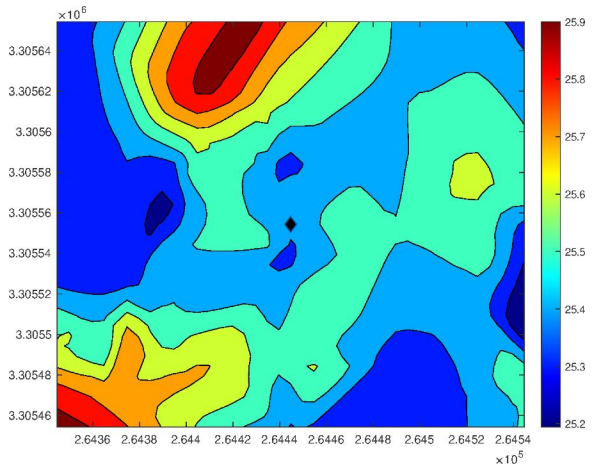

Figure 5.15: Local area contour map for the location corresponding to Figure 5.14c. The black diamond denotes the location and the total area is $200m \times 200m$

and running such a large number of surrogates are quite inefficient, even they are in the polynomial form. To guarantee the efficiency of the real-time forecast framework, we prebuilt Karhunen-Loeve expansions (see section 5.2.2) for the inundation depth field at each hour to reduce the dimensionality of QoI. Because of the strong regional correlation of the inundation depth, the dimensionality can be reduced to $M = 3$ in Eq (5.8) by using the first three dominant eigenvalues of the training sets without

losing much information (Figure 5.16). Consequently, we will train $3 \times 18 = 54$ surrogates for KL coefficients in Eq (5.8) instead of working on surrogates with every cell of each region individually. The performance of surrogates representing the KL coefficients is validated in Figure 5.17.

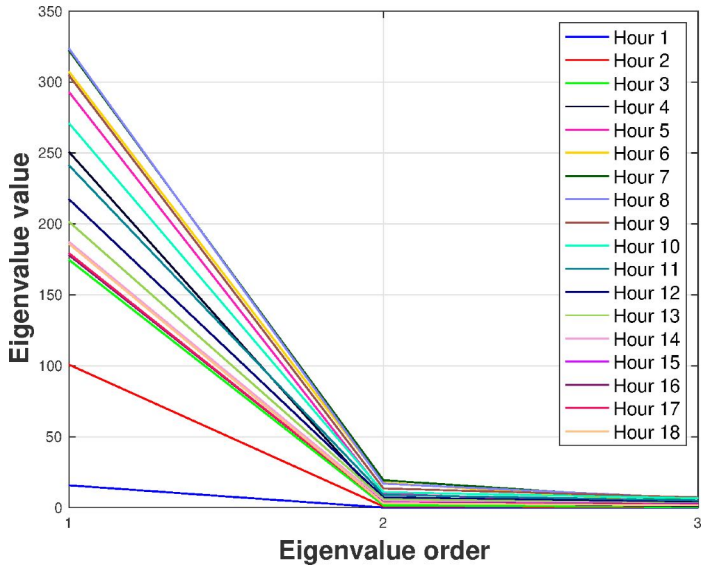

Figure 5.16: The first three eigenvalues from the training sets for Region 2 for all 18 simulation hours. Specifically, the eigenvalues are computed from $4,476 \times 1000$ matrix, where $4,476$ is the number of cells in Regions and 1000 is the number of training simulations.

5.5.4 Computational efficiency of surrogate models

Due to its polynomial form, the reduced-order surrogate model is expectedly much more computationally efficient than the high-fidelity model. About 2.84 CPU seconds (Intel(R) Xeon(R) Gold 6140 CPU @ 2.30GHz) is necessary to run all of the $127,026$ constructed surrogate models (i.e., 7507 locations 18 hours). Thus, Core Seconds per Simulated Day - Normalized (CSSD-N, Eq. (5.17)) metric for polynomial chaos surrogate (PCS) can be computed as $\eta_{PCS} = 2.84/750724/18$, or $\mathcal{O}(10^{-}4)$ CPU seconds to carry out 24-hour simulation for a single location (i.e., mesh node of the high-fidelity model). This estimate signifies 3-4 orders of magnitude difference with respect to what is typical of a high-fidelity, physical flooding model (the corresponding

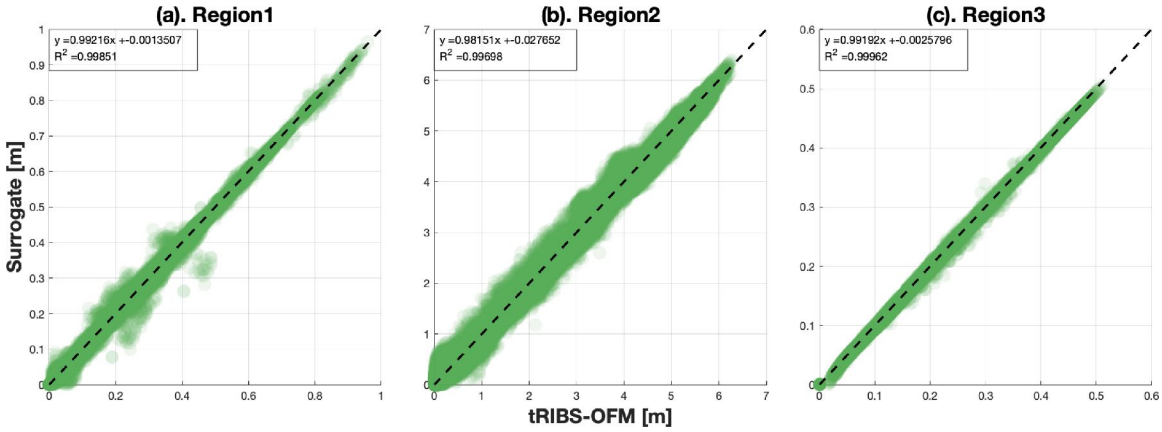

Figure 5.17: Validation of surrogates representing the KL coefficients for the inundation depth at hour 14 for all the cells in (a). Region 1, (b). Region 2, and (c). Region 3. The X-axis represents the outputs from 150 tRIBS-OFM validating simulations. The Y-axis represents the inundation depth retrieved using the KL coefficients computed from surrogates (see Eq (5.8)). The black dashed line is the reference line of $y = x$. The fitted linear regression and R^2 are inserted in the left corner of each subplot.

range is assessed as $\mathcal{O}(10^{-}1)$ - $\mathcal{O}(10^{-}0)$), which is sufficiently fast for forward runs in real-time flood-forecasting.

However, the need to carry out at the least $127,026 \times \mathcal{O}(10^{-}1) \rightarrow \mathcal{O}(10^{6})$ of basic algebraic operations embedded in polynomial expansion series per single forward run might still pose computational challenges for uncertainty analysis, especially when these are run serially on a low-performance desktop. For example, carrying out $10,000$ simulations for each QoI to quantify the uncertainty stemming from precipitation input can take about 8 hours; even more so, if the number of selected QoIs is an order(s) of magnitude higher. However, if the dimensionality of QoIs can be reduced significantly because of their high spatial correlation, there is no need to formally carry out uncertainty quantification for all QoIs. In the case study illustrated, $127,026$ QoIs can be reduced to just 54, which increases the numerical efficiency by additional 3 orders of magnitude: $\eta_{PCS} \sim \mathcal{O}(10^{-}7)$ This guarantees the efficiency of uncertainty quantification using the proposed framework relying on surrogate models.

The reallocation of intense training simulations from the flooding periods to the

time interval between them effectively replaces the strain of computational burden in real-time with a data storage problem. However the actual amount of storage is likely to remain modest: in the case study considered, 3 Gigabytes of outputs from the high-fidelity model are used to train $127,026$ surrogate models, and their multi-index arrays total in volume of ~ 1.3 Gigabytes. Normalizing the latter value by the total number of QoIs yields 0.01 Megabytes per QoI, a trivial magnitude, given the low cost of modern storage systems. Since the QoI dimensionality can be further reduced using the Karhunen-Loeve decomposition, the storage of multi-index arrays representing surrogate models is only ~ 0.5 Megabytes.

5.5.5 Sensitivity analysis

Although the PCE surrogates trained using input rainfall series of uncorrelated pulses show high skill in mimicking the tRIBS-OFM model, the performance is slightly different for QoIs in different locations (Figure 5.13). Of particular interest is Region 2, which shows obviously higher error than other regions. To investigate the reason for this, we conduct a sensitivity analysis using Eq (5.7). Figure 5.18 shows the main sensitivity Sobol index of the KL coefficient at hour 14 corresponding to the first eigenvalue, which contains over 99% information of the training sets in all the three regions. Specifically, Region 1 and Region 3 are most sensitive to the antecedent 3 hours rainfall, but up to 7 hours of previous rainfall affect Region 2 significantly. This resulted in higher variability of the inundation depths in Region 2 than other regions. However, the same number of training simulations are used to train surrogates for all three regions, which could be another reason for the relatively inferior performance of the surrogate model in Region 2.

Surrogate simulations for region 2 are sensitive to a longer period of previous rainfall, which can be attributed to several factors. First, the interaction between river and land, as Region 2 contains a segment of the river. This makes the inunda-

tion depths in Region 2 to be impacted by river depth, which integrates watershed response over a longer period. Another attribution element can be the complexity of topography and connectivity of buildings in Region 2. These two factors together influence the flow drainage route, which results in higher sensitivity.

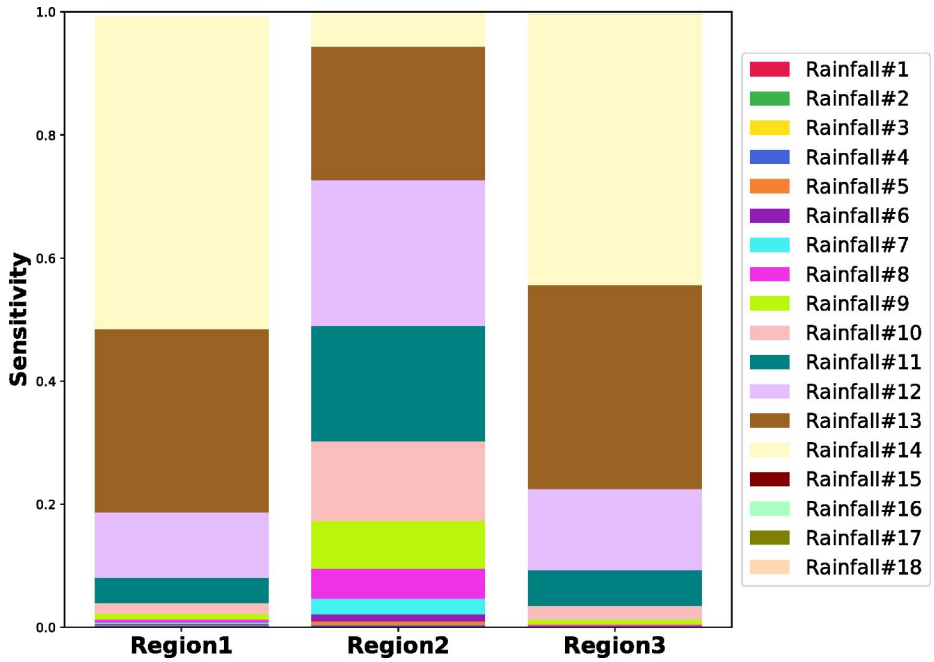

Figure 5.18: Main sensitivity analysis of rainfall to inundation depth for the three selected regions. The Y-axis represents Sobol main sensitivity index for the KL coefficients corresponding to the first eigenvalue. The sum of the Sobol indices for all the 18 hours of rainfall inputs is equal to 1.

5.5.6 Forecast with surrogates

While this work focuses only on the discharge at the outlet and inundation depths in the three regions in Figure 5.11, one needs to note that any scalar quantity in the model outputs can be selected as QoIs for constructing surrogates. Running training simulations, selecting QoIs, and constructing surrogates should be done before flood season. Then, one can use the surrogate model to predict flooding information with uncertainty quantification using the surrogate model rather than running the computationally expensive physical flooding model during the event.

The quantity precipitation estimation or QPE represented by the High-Resolution Rapid Refresh (HRRR) product (Benjamin et al., 2016) provides an 18-hour rainfall forecast in the study area. Issued at the onset of the storm caused by the Hurricane Harvey landfall, the QPE product is one of many precipitation estimates that NOAA released prior to or following this forecast. This forecast is aggregated to the watershed scale and averaged every hour since the surrogates are trained with hourly spatially uniform rainfall inputs. Further, it is used to construct a stochastic description of the rainfall process, providing uncertainty on this vital model input (Figure 5.19). By sampling from this stochastic process, a set of rainfall realizations during the forecast period is generated and used as inputs into the surrogate models of streamflow and inundation.

Specifically, Figure 5.20 illustrates the forecast of streamflow at the outlet with corresponding uncertainty bounds. While there are some differences between surrogates' prediction and USGS observations, it stems from the accuracy of the forecasted rainfall input. A snapshot of the inundation forecast of Region 1, Region2, and Region 3 is presented in Figure 5.21. The spatial distribution of uncertainty is heterogeneous; for example, the channelized area exhibits relatively larger uncertainty as compared to the land area in Region 2. Along the river stem, downstream areas exhibit higher uncertainty than upstream areas. Additionally, the area with smaller inundation depth can have higher uncertainty (Region 3, northeast corner), highlighting that forecast with uncertainty quantification is necessary. Such heterogeneity of the forecast confidence is caused by the complexity of topography and building connectivity, which can not be captured by a model that lacks spatial simulation of flooding conditions. Therefore, this framework provides a promising tool for forecasting flooding events at even a local river forecast center, which may not have access to super computation resources.

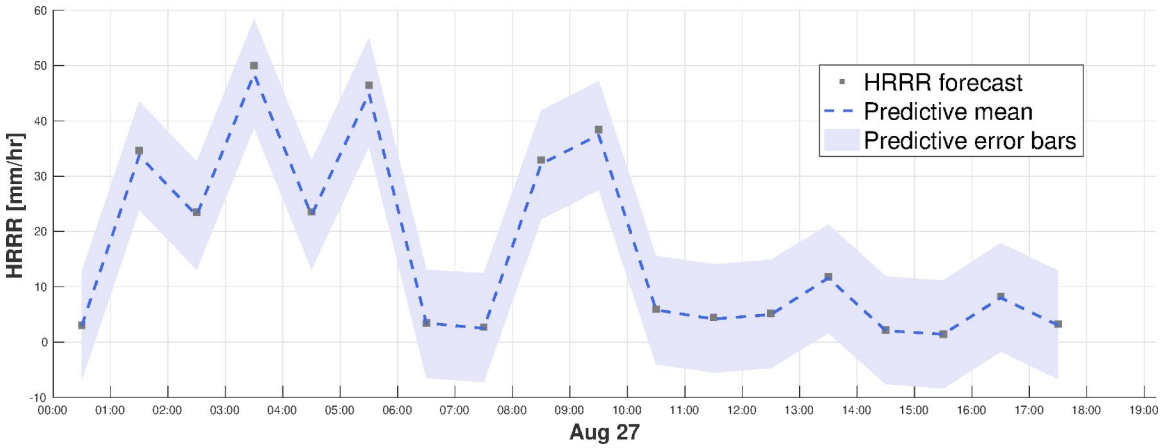

Figure 5.19: Gaussian process for the HRRR forecast. Blue dashed illustrate HRRR rainfall forecast, shaded area represents two standard deviations of the Gaussian process (GP) process fitted to the forecast rainfall series. Radial basis function kernel is used with length scale = 0.1 and variance = 533. The Gaussian noise variance is specified with 15.

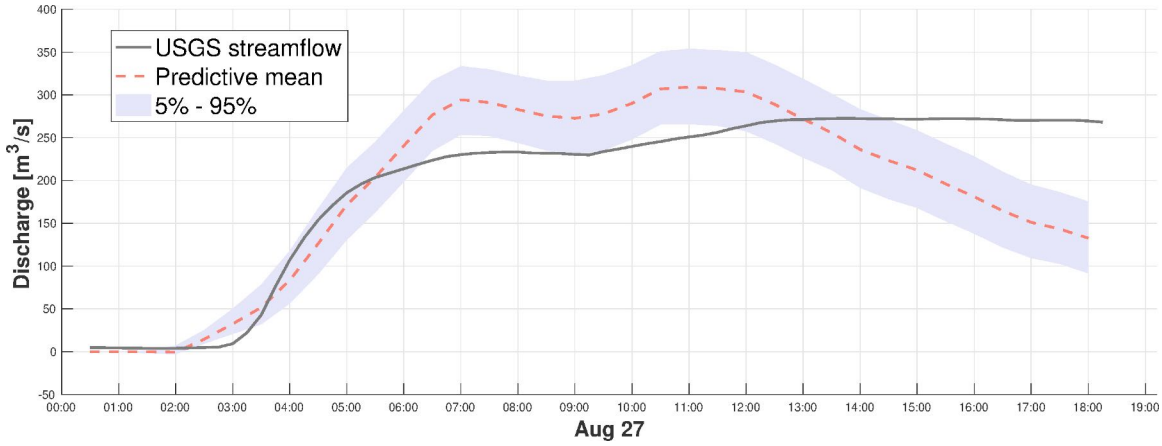

Figure 5.20: Forecast of forward 18-hour streamflow from surrogate model forced with HRRR Gaussian Process. Black solid line: the outlet discharge series from the USGS measurements. Red dashed line: the estimation results using surrogate models for 36 half-hour intervals (HRRR forecast is used as input). Blue shaded area: 5%-95% posterior probability region obtained from the surrogates that used 10,000 realizations from Gaussian Process.

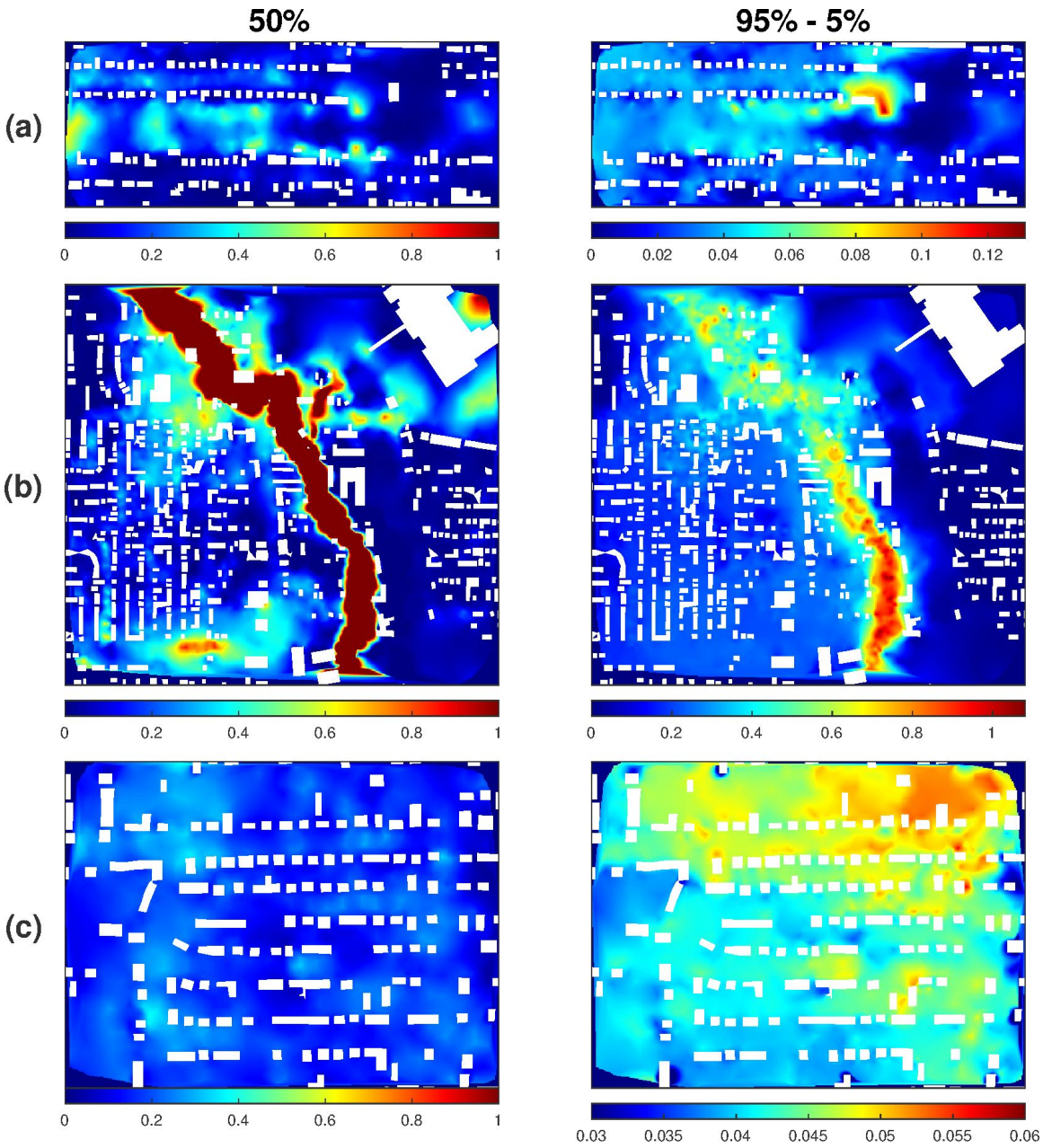

Figure 5.21: Left panel shows the inundation depth at hour 14:00 for the (a) Region 1, (b) Region 2, and (c) Region 3 estimated with surrogate models (HRRR forecast is used as input). The right panel in each row represents the uncertainty of the estimated inundation depth expressed as the depth difference between 95% and 5% of the posterior distribution obtained from the surrogates that used $10,000$.

5.6 Discussion and Conclusion

The quantity and intensity of hurricanes are projected to intensify, and this work aims to increase the ability to compute and understand risk arising from the complex processes governing flooding in urbanized areas. Specifically, in this chapter we conduct the uncertainty analysis for the urban flood simulation concerning the rainfall uncertainty. Due to the high computational burden of the flooding model that solves 2D Saint-Venant equations, we trained reduced-order models based on a polynomial chaos expansion for the physical model tRIBS-OFM. Then, the surrogate model can be used for full uncertainty quantification. While only the uncertainty of precipitation is considered in this work, one can easily include other inputs and model parameters in this framework, for example, soil moisture, roughness coefficients, hydraulic conductivity, etc.

Furthermore, a real-time urban flooding forecasting framework based on this uncertainty quantification framework has been proposed. This forecasting framework contains two steps: pre-event training and real-time forecasting (see Figure 5.2). Thousands of high fidelity flood model simulations are required to train the surrogate models in the first step. While such a physical model is computationally expensive, this can be done during no-flooding periods. Next, one can use the pre-trained surrogate model to forecast streamflow and inundation field during the flooding event. The polynomial form of the surrogate model guarantees the forecast can be finished in nearly the same time as when rainfall forecast is provided to it. Our results prove that this framework can reproduce the 2017 Harvey flooding event in a selected urbanized watershed with sufficient accuracy, and what is important, this can be achieved in real-time.

During the pre-training process, we assume that the rainfall is uniform over the whole watershed. However, Figure 5.6 illustrates that this is a simplification, and the spatial distribution is not uniform. Using the rainfall process varies in spatial for

the study domain will introduce more uncertain input variables than gauge rainfall forcing, and the number of polynomial basis terms in Eq (5.2) increases rapidly when the number of random variables grows. Consequently, many more simulations of the high-fidelity model are required to train the PCE based surrogate model, which is what we want to avoid due to limited computational resources. Additionally, applying PC expansion to problems with a high number of random variables remains challenging (*Zhang et al.*, 2017). Although the rainfall field is heterogeneous, there exists a strong regional correlation, which can be used to reduce the dimensionality of the stochastic field. For example, a two-dimensional stochastic weather generator (*Peleg et al.*, 2017) that capable of capturing the spatiotemporal evolution of the storm structure can be used to provide training rainfall series without introducing a large number of uncertain variables.

Our results show that inundation depths can exhibit threshold effects, which is caused by the complexity of topography and the effects of building connectivity on the propagation of flooding wave. This thresholding behavior of flooding propagation in the urban environment can be adequately captured by Polynomial Chaos Expansion based surrogate model, which is trained by simulation results from a distributed hydrologic and hydrodynamic model. Despite the fact that the surrogate model shows different skill in reproducing inundation depth in different regions, the overall performance is excellent.

The surrogate model can be implemented on a desktop for real-time forecasting with detailed information at high resolution due to its low computational cost. Another application of this framework is parametric inference, Specifically, given observed inundation or streamflow, one can infer the value of parameters or inputs easily through PCE theory as had been demonstrated in *Dwelle et al.* (2019). Further, this framework can be extended to future flood-smart urban design to incorporate elements of optimal drainage characteristics of the urban fabric and/or flood control.

CHAPTER VI

Research Summary and Future Efforts

6.1 Summary of research

The global temperature has already increased by $1C^{\circ}$ compared to the pre-industrial period, and this trend is projected to continue in the future (IPCC, 2018). Global warming intensifies the hydrological cycle since air can hold more moisture at a higher temperature, which will affect the life of humanity. Climate models are generally used to study the impacts of anthropogenic climate change, however, their projections are always criticized for low convergence and significant uncertainties. We have to reduce the uncertainties in some optimal fashion to get a robust conclusion of the impacts of climate change on the environment and society. Additionally, climate models' outputs cannot be used directly at local scale due to their coarse resolution. To get detailed climate information at local scale, one needs to apply the downscaling method to climate model projections. The downscaled projections can further be used as forcing for high-resolution hydraulic/hydrological/ecohydrological models to understand the response of watershed/ecosystem to climate change.

This dissertation aims to address the uncertainties in understanding hydroclimate, hydrology, and hydraulics across scales. A Bayesian Weighting Averaging (BWA) method has been coupled with an hourly weather generator in previous work to downscale hydrologic variables from climate model coarse resolution to local scale, which

represents a stochastic downscaling framework. This framework is revisited in Chapter II to downscale precipitation and temperature from 18 GCMs that participate in CMIP5 at three meteorological stations in the US. BWA method is commonly used to integrate over multi-model ensembles of climate series to extract climate change information, which will be used to estimate parameters in the weather generator to generate future meteorological series. Observations are generally thought to be relevant for correcting biases in model outputs in this process. However, they concurrently may introduce unpredictable impacts in the context of the downscaling process, in particular, when model output on precipitation is of interest. Specifically, the posterior distribution may excessively depend on few 'outlier models' being close to the observation, when all other models fail to capture observation of the historical perioda common situation for precipitation metrics. Another issue emerges for climates with very dry months: the inclusion of observation in BWA may result in a significant spread of the posterior distribution into the negative region. To address these problems, a modified version of the BWA method that removes observations in the initial phase of downscaling (computation of Factors of Change) and adds them in the estimation of posterior distributions is explored in chapter II. Comparisons of simulation results between the modified BWA and the traditional BWA demonstrate consistent outcomes with regards to the effect of observation in the Bayesian method. Further, the modified BWA approach generally reduces uncertainty, as compared to 'simple averaging' in the Bayesian context, which assigns equal weights to all model outputs.

When downscaling the hydrological variable of interest to many locations at once, a multivariate BWA (*Smith et al.*, 2009) is preferred than the univariate version used in the previous chapter. Chapter III applied the multivariate BWA method to 10 GCMs to infer the change of peak annual runoff timing under different future periods. Based on the BWA posteriors, we identified a clear spatial pattern for the shift of

peak annual runoff timing, which is driven by global warming. Specifically, the areas where runoff is dominated by snowmelt will experience an early shift maximum annual runoff. For the areas that the peak annual runoff is projected to delay, arguably, we found the centroid of wetness of soil moisture moves to a later time. This pattern tends to be stronger as the projection of more temperature increase in the future period and can result in considerable economic and environmental consequences (*Blöschl et al.*, 2017). Hence, it is urgent for us to reduce the CO_2 emissions to mitigate the impacts of climate change.

There are abundant stations/sites measuring precipitation, temperature, and streamflow, which can be used to verify and calibrate models. Those observation has been used to correct the bias of GCMs and reduce projection uncertainties. Unlike these variables, measurements of ET are sparse in space and time, and their errors are generally high 20% (*Wilson et al.*, 2002). Thus, models are required to simulate ET at any scale since it represents an important component of the surface water and energy balance that transfers water from land surface to atmosphere. Existing ET products rely on classical methods that require input data on water vapor gradient, wind speed, and surface roughness as well as model parameters that are often difficult to obtain from in situ and remote sensing observations. The uncertainty of these variables can be exceptionally large for the Amazon rainforest with high biodiversity, making it challenging to evaluate ET in space and time. In Chapter IV, a novel energy partition method based on the maximum entropy production (MEP) theory is applied to simulate ET. This method is first validated by using site-level eddy flux data from the Amazon basin and shows high skill at the hourly, daily, and monthly scales. Consistent performance under different levels of land surface dryness is revealed, hinting that drought signal is appropriately resolved. The sitelevel MEP-based estimates outperform the estimates of the Moderate Resolution Imaging Spectroradiometer evapotranspiration product, which is commonly used for largescale

assessments. At the Amazon basin scale, the two series yield similar averages but exhibit spatial differences. The parameter parsimony and demonstrated the skill of the MEP method make it an attractive approach for environments with diverse strategies of water flux control. Then we apply the MEP method to project the change of ET at the US continent scale in the context of GCM. Compare to the GCM-based estimates, the MEP-based estimates are less biased in terms of annual magnitude, and the projections of change in the future are more robust among the selected models.

Another crucial hydrological variable in the water budget is runoff, whose major phase is always related to flooding events in urban areas. A high-fertility physically-based model is required to modeling the process of runoff in the urban environments since the topography and building connectivity determine the drainage routes. Thus, Chapter V relies on tRIBS-OFMa physically rigorous hyper-resolution hydraulic and hydrologic modelto simulate flood propagation and advance numerical representation and understanding of interactions between flooding and the urban environment. Uncertainty quantification of the urban flooding simulation is necessary because of the uncertainty of the surface roughness, the difficulty of measuring or forecasting rainfall at the high resolutions and accuracies, and other parameters/inputs. However, the steep computational cost and constraints associated with resolving the 2D Saint-Venant equations at very high resolutions prevent such uncertainty analysis. We approach this problem by coupling tRIBS-OFM with an uncertainty quantification framework to pre-train simulation areas on an uncertain rainfall process arising from extreme weather events. The framework relies on polynomial chaos expansions, which creates a mapping of flooding outputs from the uncertain rainfall. This inexpensive surrogate model allows near-instantaneous predictions of quantities of interest, therefore, we turn this uncertainty quantification framework to a real-time flooding forecast framework. We justify this framework by reproducing the 2017 Havery event in an urban watershed of Houston city. Given a forecasted rainfall product, the surro-

gate model is able to predict inundation depths, streamflow, and flow velocities with uncertainty in a few seconds.

6.2 Research assumptions and limitations

Several assumptions and limitations underlie this dissertation, which can be found in the following.

6.2.1 Chapter II

- The GCM from each institution is assumed to be independent of each other. But some research argued some GCMs are dependent since they share the same boundary condition, initial condition, or same processes. The dependent GCMs may be overweighted in BWA and result in less objective inference.
- We assume that GCMs produce reasonable estimates of the change for various climate statistics, even though not necessarily accurate absolute magnitudes.

6.2.2 Chapter III

- We use the maximum runoff date for each year to represent the seasonality of the major phase of runoff. This method can result in problematic results at the regions which have multi-modal flood seasonality.
- This work cannot detect the change in peak runoff generation mechanism under future conditions. For example, the snow amount reduces significantly due to global warming in a region, where the peak runoff is caused by spring snowmelt in the past. The peak runoff generation mechanism transits to precipitation related in the future.
- We assume the shift of maximum snow amount date correlates with the shift of the maximum snowmelt date.

6.2.3 Chapter IV

- The maximum entropy production method requires only three single level inputs: net radiation, temperature, and specific humidity at the evaporating surface. For the transpiration, the temperature and specific humidity should be measured at the leaf surface, however, this level measurements are not available. In this study, we use air specific humidity (q_a) and saturated specific humidity (q_{sat}) as surrogate of leaf surface specific humidity (q_s). Using q_a results in underestimation of ET since $q_a < q_s$ based on the water vapor gradient from the leaf surface to atmosphere. The estimates of ET with q_{sat} trends to overestimate as water limitation is not included.

- We assume a linear relationship to transform Leaf Area Index (LAI) to vegetation fraction in GCM grid cell. This assumption may result in bias in estimates of ET in MEP theory since the relationship between LAI and vegetation fraction is not straightforward.

- Since there is not enough ET measurement data to validate the outputs from GCMs, we use the water budget method to estimate annual ET assuming the storage change term is negligible. The historical period we selected is 30 years, which is long enough for this assumption. However, it only applies to annual ET, and we cannot validate the GCMs' ET in terms of seasonality.

6.2.4 Chapter V

- The soil is assumed to be saturated in the simulation of tRIBS-OFM. This assumption is appropriate for the 2017 Harvey event since there is a significant volume of rainfall before our selected start time that saturated the soil. To extend this framework to a general applications, initial soil moisture and other inputs need to be considered as uncertain during the training process.

- An uncertain hourly gauge rainfall time series is accepted here to train the surrogate, which is reasonable due to the size of the study domain is small. When a larger domain is considered, the spatial structure and finer temporal resolution are necessary. However, this will introduce more number of uncertain input variables, which represents a challenge for training PCE based surrogates.

- This framework only accounts for parameter/input uncertainty, but no model structure uncertainty. For example, the building footprints are simplified before they are merged into the mesh to reduce the number of computational cells and eliminate tiny cells. While simplification improves computational efficiency, it also lowers the accuracy of the representation of the real-world in the model as a trade-off.

- The only observation that we can use to validate the model or perform inference is the streamflow from the USGS monitoring gauge. However, the streamflow during our simulation period is significantly higher than the rating limit, implying low confidence in the measurement itself. Additionally, inundation depth data are even sparse and poor in quality.

6.3 Future works

6.3.1 Transition from CMIP5 to CMIP6

Climate models simulate the physics, chemistry, and biology of the atmosphere, land, and oceans in great detail, representing an important tool for us to understand how the earth system works and its response to climate change. These models are continually being improved as more institutions, models, experiments, data, new physical processes, and biogeochemical cycles are involved. However, these improvements in models didn't lead to reduced uncertainties of projections from CMIP5 compared to

CMIP3 (*Knutti and Sedláček*, 2013). For some processes, the older version models even show better skills than the new version (*Mueller and Seneviratne*, 2014).

The next phase of Coupled Model Intercomparison Projects (CMIP) is CMIP6, which will be featured in the upcoming 2021 IPCC sixth assessment report (AR6). A lot of CMIP6 models are available to download and use to the date of this dissertation, and more models will continue to release in the future (see Table 6.1). CIMP6 is a substantial expansion over CMIP5 with respect to the number of models. Another significant difference is that CMIP6 provides more reasonable future scenarios to project how policies affect greenhouse gas emissions in the future. Hence, we should transit to CMIP6 for future studies. The first question is if the great efforts of updating CMIP6 models will lead to better performance in reproducing climate series in the historical periods and more convergence in projecting climate change in future periods?

For example, our results in Chapter IV show that CMIP5 models overestimate ET in the western of the US, which correlates with the overestimation of precipitation. Unfortunately, this bias pattern still exists in a similar pattern, as shown in Figure 6.1. Evaluation of other variables and in-depth analysis should be implemented in future studies. But one needs to note, more sophisticated process, higher spatial resolution, and more calibrating data does not necessarily narrow the uncertainty, as these efforts are added to represent a more comprehensive model (*Knutti and Sedláček*, 2013). While the uncertainty may remain unchanged, there is more confidence in the projections due to the additional efforts.

6.3.2 High resolution ET for Amazon rainforest

Our results in Chapter IV illustrate that the estimation of ET is not sensitive to mild droughts in the Amazon rainforest, while water limitation constrains ET by closing stomata in some extent. There are two possibles explanations: 1) the plants

Table 6.1: A list of CMIP6 models

No.	Model name	No.	Model name
1	CESM2-FV2	17	INM-CM4-8
2	CESM2-WACCM	18	INM-CM5-0
3	CESM2-WACCM-FV2	19	IPSL-CM6A-LR
4	CanESM5	20	KACE-1-0-G
5	E3SM-1-0	21	MCM-UA-1-0
6	E3SM-1-1	22	MIROC6
7	E3SM-1-1-ECA	23	MPI-ESM-1-2-HAM
8	EC-Earth3	24	MPI-ESM1-2-HR
9	EC-Earth3-Veg	25	MPI-ESM1-2-LR
10	FGOALS-f3-L	26	MRI-ESM2-0
11	FGOALS-g3	27	NESM3
12	FIO-ESM-2-0	28	NorCPM1
13	GFDL-ESM4	29	NorESM2-LM
14	GISS-E2-1-G	30	NorESM2-MM
15	GISS-E2-1-G-CC	31	SAM0-UNICON
16	GISS-E2-1-H	32	TaiESM1

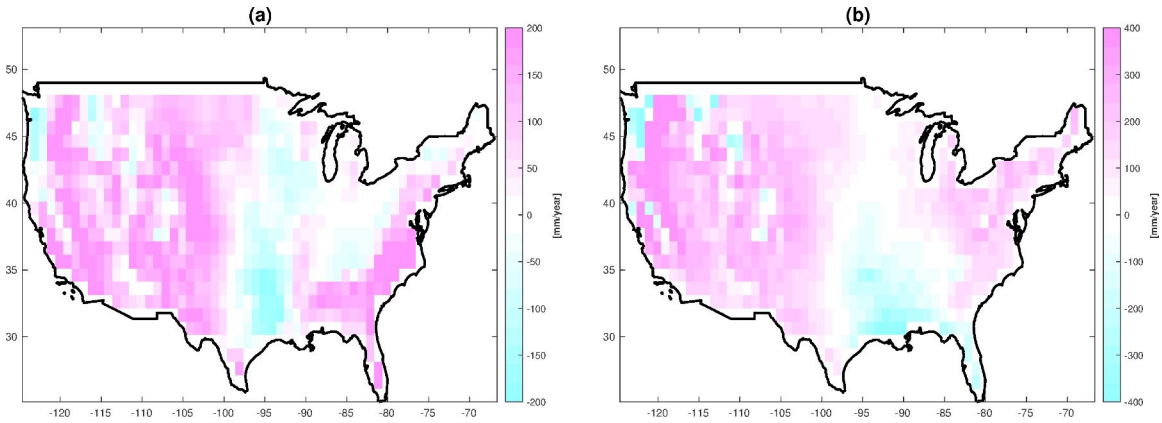

Figure 6.1: Spatial difference between *Livneh et al.* (2013) and multi-model mean of CMIP6 in terms of averaged (a) annual ET, and (b) annual precipitation in the control period (1961-1990). The multi-model mean of CMIP6 is computed with all the GCMs listed in Table 6.1 with equal weights.

are resilient to the seasonal water limitations and the mild ENSO driven drought events (e.g., 2005 and 2010) through up-taking water from deep soils; 2) the sites we used in our analysis are not affected by droughts. The future direction is to extend the analysis of ET to the whole basin and longer period including 2015 drought, which is considered to be more severe than the previous droughts. This can be done by

driving MEP with satellite data. For example, the following products provide all the required inputs for the MEP model:

1. NASA's GOES Surface and Insolation Product (GSIP) version 3 product provides net radiation at hourly temporal resolution and 4 km spatial resolution.

2. Hourly specific humidity, surface temperature, vegetation fraction, and soil moisture are taken from ERA5, which is the fifth generation of reanalysis product produced by the European Centre for Medium-Range Weather Forecasts (ECMWF). The spatial resolution of this product is $0.25° \times 0.25°$.

Figure 6.2 illustrates an example of MEP based ET driven by above forcings using the transpiration version formulation (Eq (4.10)) and assume the ground heat flux is negligible. This assumption is appropriate for most of the Amazon basin but can result in significant overestimation of ET in the southeast of the basin, where the land cover is savannas, grasslands, croplands, etc (see Figure 4.13). Therefore, a full version of MEP (Eq (4.15)) should be used to provide more accurate ET at the basin scale. Another problem we need to address in the future is that which depth of soil moisture represents the water supply in the ET process? While *Hajji et al.* (2018) proposed a relationship between top layer soil and water limitation to improve the performance of MEP theory under significant water stress, it doesn't apply to the Amazon rainforest because of the deep root systems (*Ivanov et al.*, 2012).

The high-resolution ET allows us to investigate 1). what is the threshold of drought severity that will affect the plant functionality; 2). which part of the Amazon rainforest is more vulnerable to drought, and 3) legacy effects of 2015 mega drought to the basin.

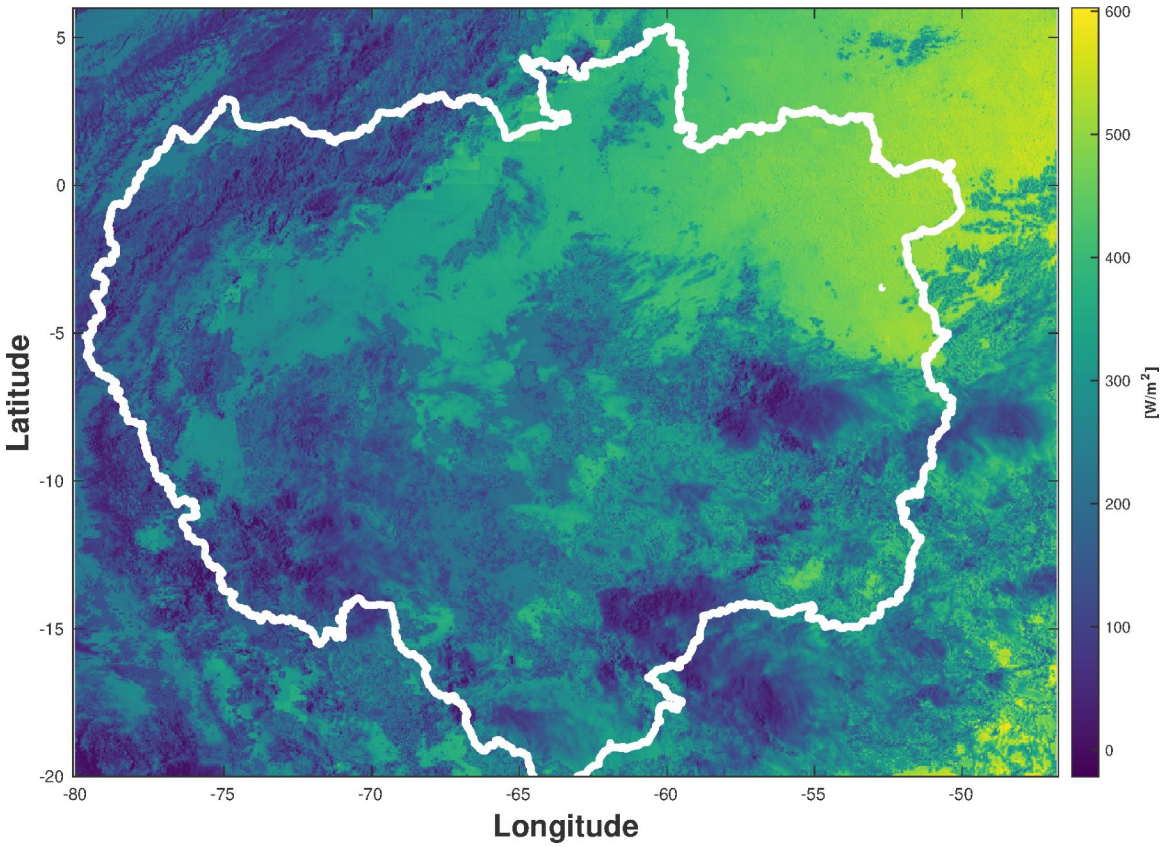

Figure 6.2: An example of MEP based ET driven by GSIP3 net radiation, and ERA5 specific humidity and surface temperature using Eq (4.10). The spatial resolution is $4kmkm$ and the time associate with this snapshot is 03/20/2016 12:45pm.

6.3.3 Deep learning of tRIBS-OFM for real-time flood forecasting

We train a surrogate model of the physical flooding model (e.g., tRIBS-OFM) with Polynomial Chaos Expansion (PCE) in Chapter V. The PCE based surrogate model shows high skill in reproducing the outputs of tRIBS-OFM, such as inundation depths, discharge, and flow velocity. However, the performance is not consistent for all the selected Quantity of Interests. For example, Figure 5.14 illustrates some Quantity of Interests that the accuracy of the surrogate model is lower than others because of high non-linearity. While the overall performance of PCE based surrogate is excellent, the next question is if we can improve the performance of surrogate models for those high non-linear Quantity of Interests? The PCE is truncated at order 3 during the

training process, but higher order doesn't improve its performance (not shown here). Thus, we need to consider other methods.

Neural networks can learn high non-linear function using layers of neurons which mathematically transform the data and is computationally cheap, which provides an alternative option for the surrogate model. The inundation depths of Region 2 at hour 14 from 1,000 training simulations in Chapter V are used to train a fully-connected ReLU network with two hidden layers presented in Figure 6.3 in this section. Specifically, this neural network contains an input layer for the 18 rainfall inputs, 2 hidden layers with 8 neurons in each one, and a single neuron in the output layer that represents the scalar Quantity of Interest from flooding model. 150 validating simulations are used to validate the performance of this neural network and compare it with the PCE surrogate model that we used in Chapter V. Figure 6.4a shows the comparison of PCE and neural network in reproducing the KL coefficients of the first eigenvalue of KL decomposition (see Eq (5.8)), which represents the overall performance for all the inundation depths in Region 2. The averaged errors using Eq (5.12) for the validating simulation (training simulation) are **0.32** (**0.19**) and **0.22** (**0.09**) for PCE and neural network, respectively. Subplot (b) illustrates the comparison corresponding to inundation depths from one Quantity of Interest that PCE cannot capture accurately, and the relative error is 0.25. The performance of the surrogate model is improved by using the neural network for this specified Quantity of Interest with relative error reduced to 0.16. In summary, the performance of a simple neural network is better than Polynomial Chaos Expansion, while some non-linear behavior remains poorly resolved. In the future, other architectures and activation functions should be tried.

6.3.4 Impacts of climate change on urban flooding

Changes in flooding can be due to individual or combined changes in precipitation, land use, and land cover, as well as practices of water management. These changes

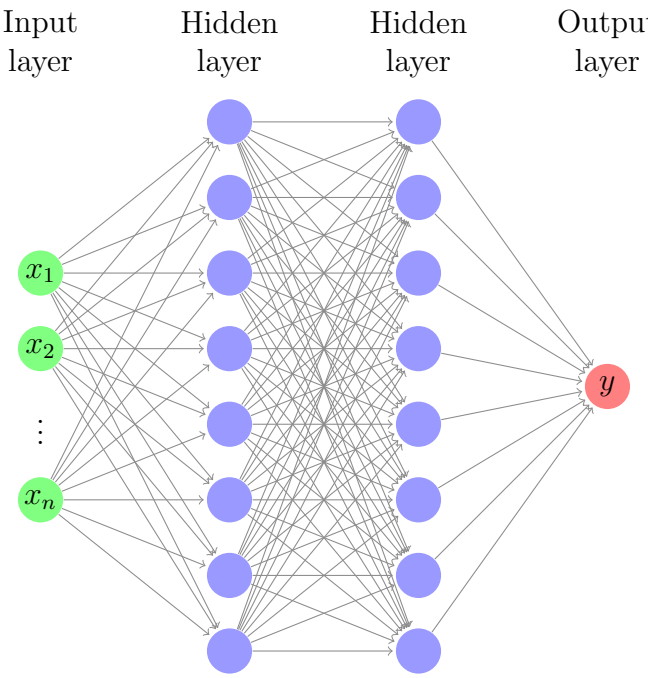

Figure 6.3: Neural Network used in this study. 18 hourly rainfall are used as inputs, and two hidden layers are selected with 8 neurons for each. There is a scalar in the output layer, which is the QoI.

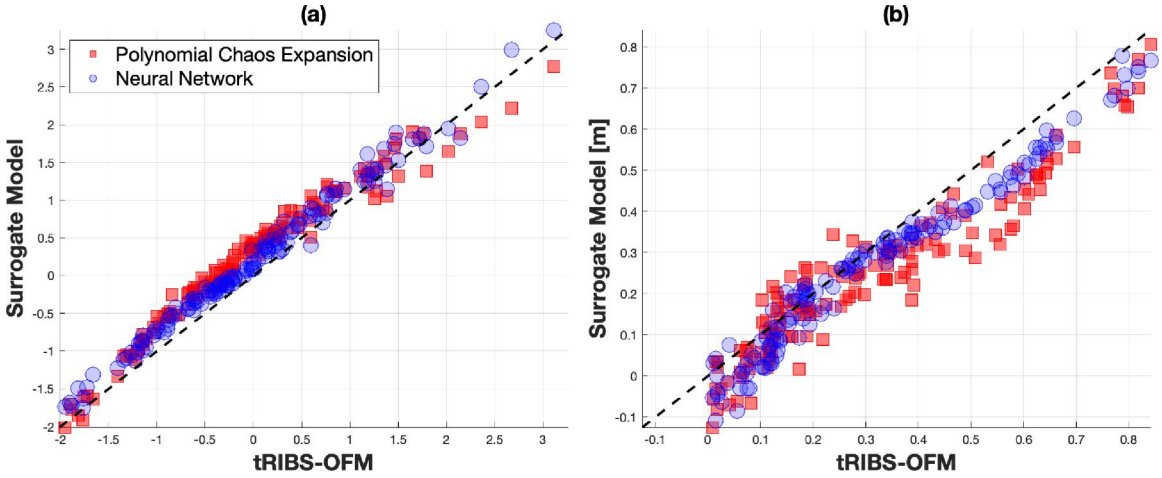

Figure 6.4: Comparison of performance of Polynomial Chaos Expansion and Neural Network in reproducing 150 validating simulations for (a) the KL coefficients of first eigenvalue derived from inundation depth of Regions 2 at hour 14, and (b) inundation depth at hour 14 from an individual cell in Region 2 corresponding to higher error. In both subplots, X-axis represents results from tRIBS-OFM and Y-axis represents results from surrogate model.

can co-occur and make it difficult to tease out the relative importance of each factor as drivers of observed changes in river flooding behavior. Given empirical evidence,

however, it is clear that the hydroclimate has already changed, and changes will cascade into the watershed response. For example, Chapter III shows a clear pattern of change of peak annual runoff timing induced by global warming, which is always associated with flooding events. Despite a lot of literature discussed the change of flooding characteristics (see Chapter III), there is a lack of such evaluation at "human action" scale (e.g., $\mathcal{O}(10^0) \sim \mathcal{O}(10^1) \quad meter$). There are two major challenges for such analysis: 1) Global Climate Models don't simulate extreme events well, and 2) computational burdens of flooding model in urban environments make is hard for frequency analysis with long term period simulations.

These two challenges can be addressed by the frameworks/methods implemented in this dissertation. Specifically, the stochastic downscaling framework we investigated in Chapter II is able to downscale the climate forcings (e.g., temperature and precipitation) from the coarse GCM resolution to local scale and capture the stochastic variability and low-frequency properties of variables. Figure 6.5 shows the change of return period in the end-century with RCP85 scenario in terms of annual maximum precipitation at 7 WebMET stations in Michigan. Specifically, hourly rainfall realizations of 3,000 years are used for both periods (historical and future) to capture the natural variability and any possible extreme events. Then, the uncertainty quantification framework we present in Chapter V provides an efficient tool to convert the change of precipitation to the change of runoff/streamflow.

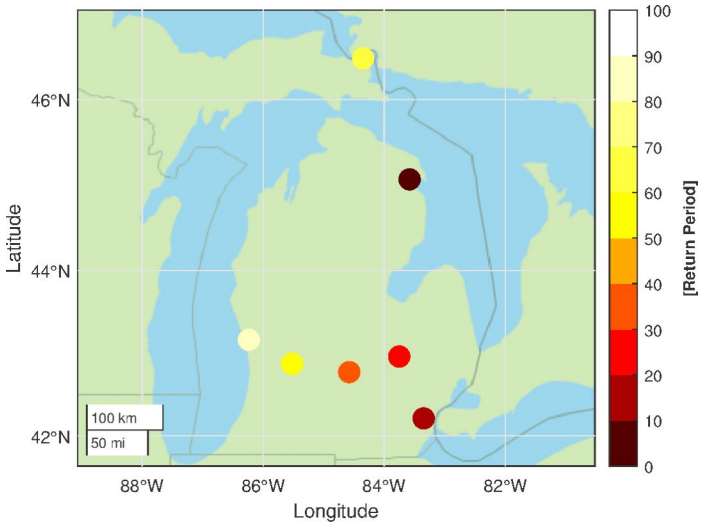

Figure 6.5: Projected change in annual maximum precipitation frequency over seven WebMET locations in Michigan. It shows return period (years) in 2071-2100 under RCP85 for annual maximum precipitation corresponding to the 100-year return period under the control period (1961-1990) condition.Stochastic downscale framework presented in Chapter II is used to downscale the precipitation projections from 18 GCMs to the seven selected locations. 3,000 year hourly precipitation is generated from a Weather Generator (*Fatichi et al.*, 2011) for both the control period (1961-1990) and future period (2071-2100, RCP85).

APPENDICES

APPENDIX A

BWA Result Comparison of All Cases for Precipitation and Temperature

A comparison of results for all the simulations scenarios can be found in Figure A.1 and Figure A.2 and one can note that they are consistent with the results presented in Chapter II: (1). FOC distributions corresponding to traditional BWA (*Tebaldi et al.*, 2005) and revised BWA are quite similar, except for some months when observations are not captured by the GCM output ensemble; (2). Modified BWA has better results for dry months, such as the month of June for the location of Fresno; (3). The difference between temperature FOCs based on two methods is marginal for all cases.

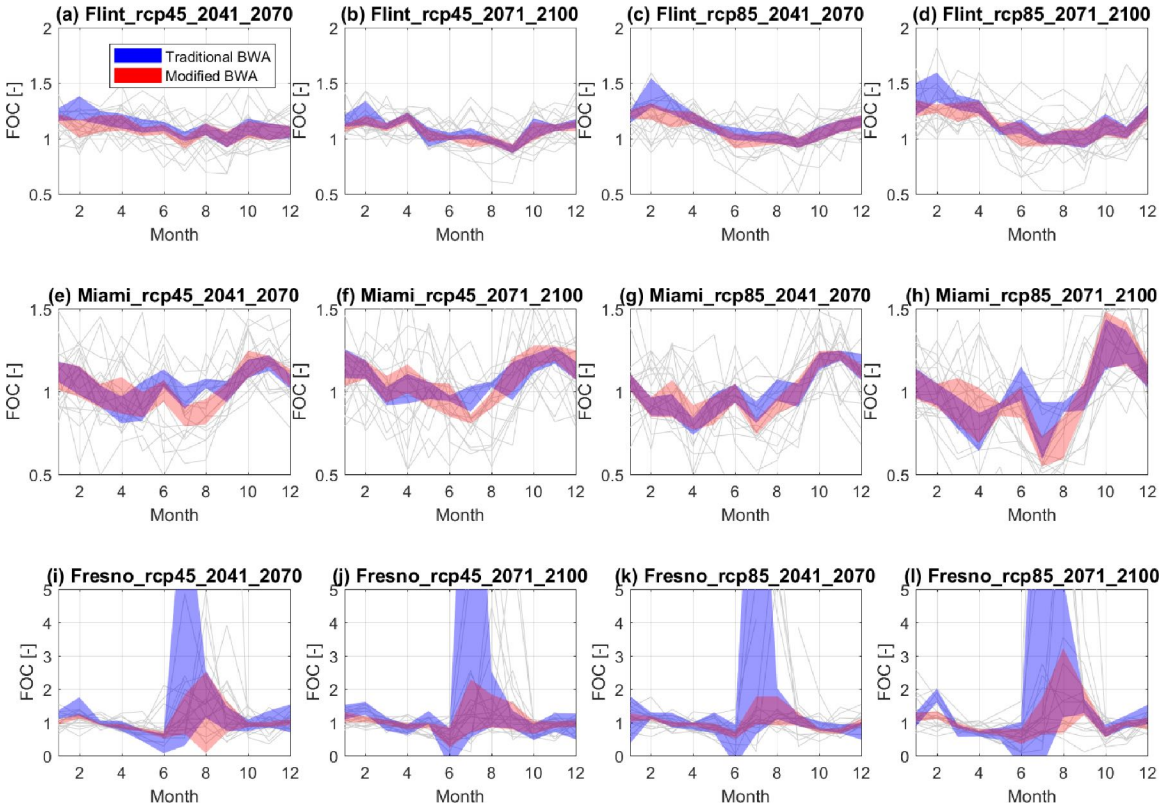

Figure A.1: The 5th - 95th percentile uncertainty bounds of the posterior distribution of FOCs of precipitation for all considered cases, including 3 locations: Miami, Fresno, and Flint; 2 emission scenarios: RCP45 and RCP85; and 2 time windows: mid-century (2041-2070) and end-century (2071-2100). The light grey solid lines represent FOCs of monthly precipitation from individual GCM.

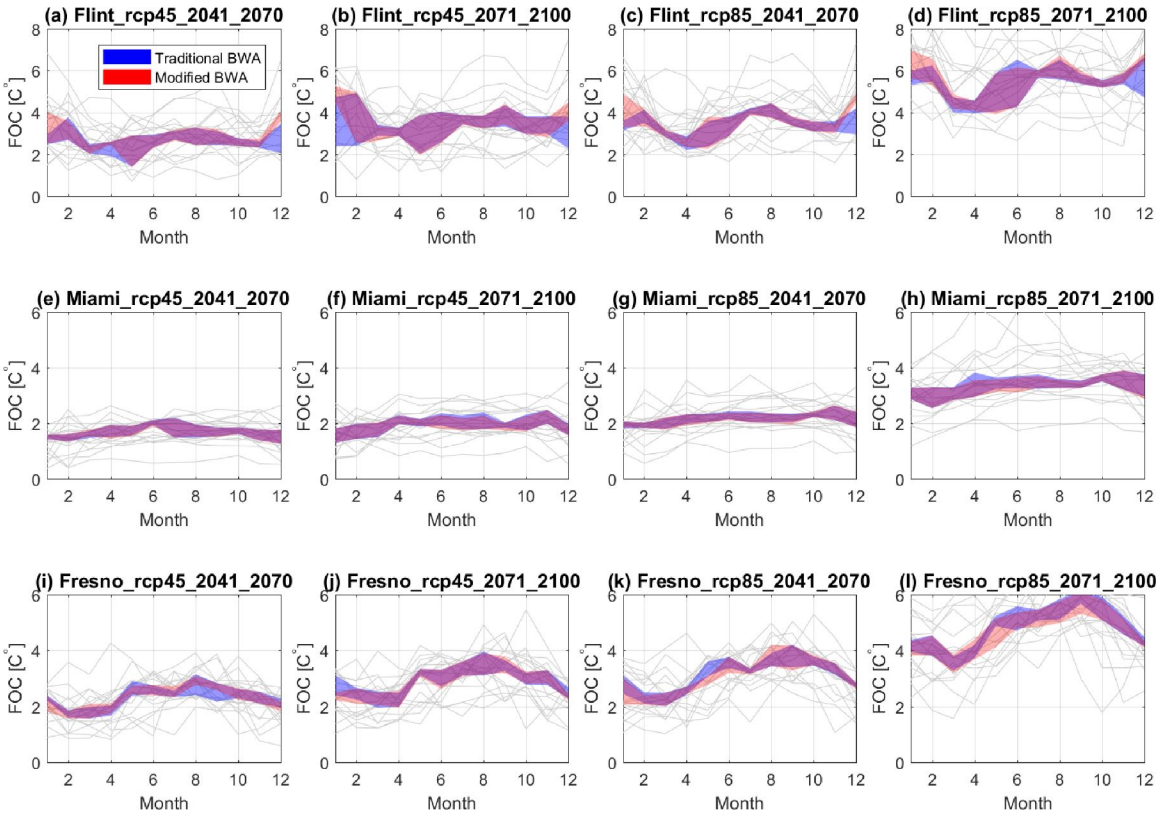

Figure A.2: The 5th - 95th percentile uncertainty bounds of the posterior distribution of FOCs of temperautre for all considered cases, including 3 locations: Miami, Fresno, and Flint; 2 emission scenarios: RCP45 and RCP85; and 2 time windows: mid-century (2041-2070) and end-century (2071-2100). The light grey solid lines represent FOCs of monthly temperature from individual GCM.

APPENDIX B

Validation of precipitation of *Livneh et al.* (2013)

The mean annual precipitation of *Livneh et al.* (2013) between 1961 and 1990 are validated with observed precipitation from 183 WebMET (`www.webmet.com`) stations, whose locations can be found in Figure B.1b. Figure B.1a confirms that the mean annual precipitation of *Livneh et al.* (2013) agrees well with observation for the selected period overall as the R^2 is pretty high and the slope of the regression line approaches 1. However, *Livneh et al.* (2013) overestimates the annual precipitation at about 13% of the WebMET stations significantly (relatively error $> 10\%$).

Although there is sparse observations of ET, one can get a reliable annual estimates in longer term through water budget method: $ET = P - R$. The change of water storage is generally negligible in long term. The accuracy of water budget method is high since the precipitation and runoff are verified by observation with good performance. Because the precipitation product overestimate the observation (Figure B.1a), and the runoff product underestimate the observation slightly (Figure 3.3, the water budget method is expected to yield higher estimate of ET.

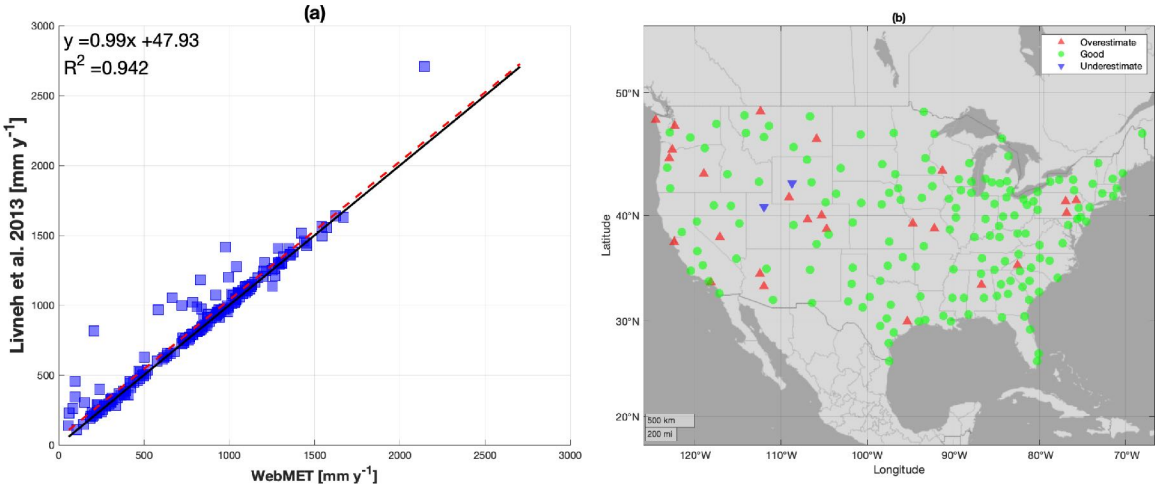

Figure B.1: Validation of annual mean precipitation of *Livneh et al.* (2013) during 1961-1990 with WebMET observation. Subplot a shows the scatter plot with regression line (red dashed line) of the comparison, and the black solid line represent the 1:1 reference line. The spatial location of the WebMET locations can be found in subplot b, the red triangles and blue triangles indicate the relative error is larger than 10% with positive and negative bias, respectively. The green circle markers represent good agreement (relative error less than 10%) between the product of *Livneh et al.* (2013) and WebMET observation

APPENDIX C

Derivatives for Manufactured Solution

The derivatives of manufactured solution (Eq 5.15) that appear in the governing equation 5.13 of 2D Saint-Venant can be evaluated exactly as following:

$$\frac{\partial h}{\partial t} = \frac{h_0}{t_0}(1 + \sin(\frac{\pi x}{L_x})\sin(\frac{\pi y}{L_y}))\exp(t/t_0)$$

$$\frac{\partial uh}{\partial t} = u\frac{\partial h}{\partial t} + h\frac{\partial u}{\partial t} = \frac{2u_0h_0}{t_0}\left[(1 + \sin(\frac{\pi x}{L_x})\sin(\frac{\pi y}{L_y}))\exp(t/t_0)\right]^2$$

$$\frac{\partial vh}{\partial t} = v\frac{\partial h}{\partial t} + h\frac{\partial v}{\partial t} = \frac{2v_0h_0}{t_0}\left[(1 + \sin(\frac{\pi x}{L_x})\sin(\frac{\pi y}{L_y}))\exp(t/t_0)\right]^2$$

$$\frac{\partial uh}{\partial x} = u\frac{\partial h}{\partial x} + h\frac{\partial u}{\partial x} = \frac{2\pi u_0 h_0}{L_x}\left[(1+\sin(\frac{\pi x}{L_x})\sin(\frac{\pi y}{L_y}))\exp(t/t_0)\right]\left[\cos(\frac{\pi x}{L_x})\sin(\frac{\pi y}{L_y})\exp(t/t_0)\right]$$

$$\frac{\partial u^2 h + \frac{1}{2}gh^2}{\partial x} = h\frac{\partial u^2}{\partial x} + u^2\frac{\partial h}{\partial x} + gh\frac{\partial h}{\partial x} = 2uh\frac{\partial u}{\partial x} + (u^2+gh)\frac{\partial h}{\partial x}$$

$$= 2u_0 h_0\left[(1+\sin(\frac{\pi x}{L_x})\sin(\frac{\pi y}{L_y}))\exp(t/t_0)\right]^2\frac{\pi u_0}{L_x}\left[\cos(\frac{\pi x}{L_x})\sin(\frac{\pi y}{L_y})\exp(t/t_0)\right] +$$

$$\left[u_0^2\left[(1+\sin(\frac{\pi x}{L_x})\sin(\frac{\pi y}{L_y}))\exp(t/t_0)\right]^2 + gh_0(1+\sin(\frac{\pi x}{L_x})\sin(\frac{\pi y}{L_y}))\exp(t/t_0)\right]\times$$

$$\frac{\pi h_0}{L_x}\left[\cos(\frac{\pi x}{L_x})\sin(\frac{\pi y}{L_y})\exp(t/t_0)\right]$$

$$\frac{\partial uvh}{\partial x} = uh\frac{\partial v}{\partial x} + vh\frac{\partial u}{\partial x} + uv\frac{\partial h}{\partial x} = \frac{3\pi u_0 v_0 h_0}{L_x}\left[(1+\sin(\frac{\pi x}{L_x})\sin(\frac{\pi y}{L_y}))\exp(t/t_0)\right]^2\times$$

$$\left[\cos(\frac{\pi x}{L_x})\sin(\frac{\pi y}{L_y})\exp(t/t_0)\right]$$

$$\frac{\partial vh}{\partial y} = v\frac{\partial h}{\partial y} + h\frac{\partial v}{\partial y} = \frac{2\pi v_0 h_0}{L_y}\left[(1+\sin(\frac{\pi x}{L_x})\sin(\frac{\pi y}{L_y}))\exp(t/t_0)\right]\left[\sin(\frac{\pi x}{L_x})\cos(\frac{\pi y}{L_y})\exp(t/t_0)\right]$$

$$\frac{\partial uvh}{\partial y} = uh\frac{\partial v}{\partial y} + vh\frac{\partial u}{\partial y} + uv\frac{\partial h}{\partial y} = \frac{3\pi u_0 v_0 h_0}{L_y}\left[(1+\sin(\frac{\pi x}{L_x})\sin(\frac{\pi y}{L_y}))\exp(t/t_0)\right]^2\times$$

$$\left[\sin(\frac{\pi x}{L_x})\cos(\frac{\pi y}{L_y})\exp(t/t_0)\right]$$

$$\frac{\partial v^2 h + \frac{1}{2}gh^2}{\partial y} = h\frac{\partial v^2}{\partial y} + v^2\frac{\partial h}{\partial y} + gh\frac{\partial h}{\partial y} = 2vh\frac{\partial v}{\partial y} + (v^2+gh)\frac{\partial h}{\partial y}$$

$$= 2v_0 h_0\left[(1+\sin(\frac{\pi x}{L_x})\sin(\frac{\pi y}{L_y}))\exp(t/t_0)\right]^2\frac{\pi v_0}{L_y}\left[\sin(\frac{\pi x}{L_x})\cos(\frac{\pi y}{L_y})\exp(t/t_0)\right] +$$

$$\left[v_0^2\left[(1+\sin(\frac{\pi x}{L_x})\sin(\frac{\pi y}{L_y}))\exp(t/t_0)\right]^2 + gh_0(1+\sin(\frac{\pi x}{L_x})\sin(\frac{\pi y}{L_y}))\exp(t/t_0)\right]\times$$

$$\frac{\pi h_0}{L_y}\left[\sin(\frac{\pi x}{L_x})\cos(\frac{\pi y}{L_y})\exp(t/t_0)\right]$$

BIBLIOGRAPHY

BIBLIOGRAPHY

Abe, S. (2010), *Support Vector Machines for Pattern Classification*, 2nd ed., Springer Publishing Company, Incorporated.

Abhyankar, S., J. Brown, E. M. Constantinescu, D. Ghosh, B. F. Smith, and H. Zhang (2018), Petsc/ts: A modern scalable ode/dae solver library, *arXiv preprint arXiv:1806.01437*.

Abramowitz, G. (2010), Model independence in multi-model ensemble prediction.

Abramowitz, M. (1974), *Handbook of Mathematical Functions, With Formulas, Graphs, and Mathematical Tables,*, Dover Publications, Inc., USA.

Akhtar, M., N. Ahmad, and M. Booij (2008), The impact of climate change on the water resources of hindukushkarakorumhimalaya region under different glacier coverage scenarios, *Journal of Hydrology*, *355*(1), 148 – 163, doi:https://doi.org/10.1016/j.jhydrol.2008.03.015.

Allan, R. P., C. Liu, M. Zahn, D. A. Lavers, E. Koukouvagias, and A. Bodas-Salcedo (2014), Physically consistent responses of the global atmospheric hydrological cycle in models and observations, *Surveys in Geophysics*, *35*(3), 533–552, doi:10.1007/s10712-012-9213-z.

Anandhi, A., A. Frei, D. C. Pierson, E. M. Schneiderman, M. S. Zion, D. Lounsbury, and A. H. Matonse (2011), Examination of change factor methodologies for climate change impact assessment, *Water Resources Research*, *47*(3), doi:10.1029/2010WR009104.

Aragão, L. E. O., Y. E. Shimabukuro, F. D. E. Santo, and M. Williams (2005), Landscape pattern and spatial variability of leaf area index in eastern amazonia, *Forest Ecology and Management*, *211*(3), 240 – 256, doi:https://doi.org/10.1016/j.foreco.2005.02.062.

Aragão, L. E. O. C., Y. Malhi, R. M. Roman-Cuesta, S. Saatchi, L. O. Anderson, and Y. E. Shimabukuro (2007), Spatial patterns and fire response of recent amazonian droughts, *Geophysical Research Letters*, *34*(7), doi:10.1029/2006GL028946.

Artichowicz, W., and D. Gsiorowski (2019), Computationally efficient solution of a 2d diffusive wave equation used for flood inundation problems, *Water*, *11*(10), doi:10.3390/w11102195.

Balaji, V., et al. (2017), Cpmip: measurements of real computational performance of earth system models in cmip6, *Geoscientific Model Development*, *10*(1), 19–34, doi:10.5194/gmd-10-19-2017.

Balay, S., et al. (2019), PETSc Web page, `https://www.mcs.anl.gov/petsc`.

Baldocchi, D., et al. (2001), Fluxnet: A new tool to study the temporal and spatial variability of ecosystem-scale carbon dioxide, water vapor, and energy flux densities, *Bulletin of the American Meteorological Society*, *82*(11), 2415–2434, doi:10.1175/1520-0477(2001)082⟨2415:FANTTS⟩2.3.CO;2.

Bates, P. D. (2012), Integrating remote sensing data with flood inundation models: how far have we got?, *Hydrological Processes*, *26*(16), 2515–2521, doi:10.1002/hyp.9374.

Begnudelli, L., and B. F. Sanders (2006), Unstructured grid finite-volume algorithm for shallow-water flow and scalar transport with wetting and drying, *Journal of Hydraulic Engineering*, *132*(4), 371–384, doi:10.1061/(ASCE)0733-9429(2006)132:4(371).

Benjamin, S. G., J. M. Brown, G. Brunet, P. Lynch, K. Saito, and T. W. Schlatter (2018), 100 years of progress in forecasting and nwp applications, *Meteorological Monographs*, *59*, 13.1–13.67, doi:10.1175/AMSMONOGRAPHS-D-18-0020.1.

Bentsen, M., et al. (2013), The norwegian earth system model, noresm1-m part 1: Description and basic evaluation of the physical climate, *Geoscientific Model Development*, *6*(3), 687–720, doi:10.5194/gmd-6-687-2013.

Berens, P. (2009), Circstat: A matlab toolbox for circular statistics, *Journal of Statistical Software, Articles*, *31*(10), 1–21, doi:10.18637/jss.v031.i10.

Berghuijs, W. R., R. A. Woods, C. J. Hutton, and M. Sivapalan (2016), Dominant flood generating mechanisms across the united states, *Geophysical Research Letters*, *43*(9), 4382–4390, doi:10.1002/2016GL068070.

Beven, K. (1993), Prophecy, reality and uncertainty in distributed hydrological modelling, *Advances in Water Resources*, *16*(1), 41 – 51, doi:https://doi.org/10.1016/0309-1708(93)90028-E, research Perspectives in Hydrology.

Bi, D., et al. (2013), The ACCESS coupled model: description, control climate and evaluation, *AUSTRALIAN METEOROLOGICAL AND OCEANOGRAPHIC JOURNAL*, *63*(1), 41–64, doi:{10.22499/2.6301.004}.

Bierkens, M. F. P., et al. (2015), Hyper-resolution global hydrological modelling: what is next?, *Hydrological Processes*, *29*(2), 310–320, doi:10.1002/hyp.10391.

Bishop, C. H., and G. Abramowitz (2013), Climate model dependence and the replicate earth paradigm, *Climate Dynamics*, *41*(3), 885–900, doi:10.1007/s00382-012-1610-y.

Biudes, M. S., G. L. Vourlitis, N. G. Machado, P. H. Z. de Arruda, G. A. R. Neves, F. de Almeida Lobo, C. M. U. Neale, and J. de Souza Nogueira (2015), Patterns of energy exchange for tropical ecosystems across a climate gradient in mato grosso, brazil, *Agricultural and Forest Meteorology*, *202*, 112 – 124, doi:https://doi.org/10.1016/j.agrformet.2014.12.008.

Blöschl, G., et al. (2017), Changing climate shifts timing of european floods, *Science*, *357*(6351), 588–590, doi:10.1126/science.aan2506.

Bradford, S. F., and B. F. Sanders (2002), Finite-volume model for shallow-water flooding of arbitrary topography, *Journal of Hydraulic Engineering*, *128*(3), 289–298, doi:10.1061/(ASCE)0733-9429(2002)128:3(289).

Brum, M., et al. (2019), Hydrological niche segregation defines forest structure and drought tolerance strategies in a seasonal amazon forest, *Journal of Ecology*, *107*(1), 318–333, doi:10.1111/1365-2745.13022.

Brutsaert, W. (1982), *Evaporation into the Atmosphere: Theory, History and Applications*, Environmental Fluid Mechanics, Springer Netherlands.

Buda, A. (2013), *Surface-Runoff Generation and Forms of Overland Flow*, vol. 7, pp. 73–84, Elsevier Inc., United States, doi:10.1016/B978-0-12-374739-6.00151-2.

Burde, G. I., and A. Zangvil (2001), The estimation of regional precipitation recycling. part ii: A new recycling model, *Journal of Climate*, *14*(12), 2509–2527, doi:10.1175/1520-0442(2001)014⟨2509:TEORPR⟩2.0.CO;2.

Burlando, P., and R. Rosso (2002), Effects of transient climate change on basin hydrology. 1. precipitation scenarios for the arno river, central italy, *Hydrological Processes*, *16*(6), 1151–1175, doi:10.1002/hyp.1055.

Castro, C. L., R. A. Pielke Sr., and G. Leoncini (2005), Dynamical downscaling: Assessment of value retained and added using the regional atmospheric modeling system (rams), *Journal of Geophysical Research: Atmospheres*, *110*(D5), doi:10.1029/2004JD004721.

Chadwick, R., I. Boutle, and G. Martin (2013), Spatial patterns of precipitation change in cmip5: Why the rich do not get richer in the tropics, *Journal of Climate*, *26*(11), 3803–3822, doi:10.1175/JCLI-D-12-00543.1.

Chandler, R. E. (2013), Exploiting strength, discounting weakness: combining information from multiple climate simulators, *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *371*(1991), 20120,388, doi:10.1098/rsta.2012.0388.

Chen, J., F. P. Brissette, and R. Leconte (2011), Uncertainty of downscaling method in quantifying the impact of climate change on hydrology, *Journal of Hydrology*, *401*(3), 190 – 202, doi:https://doi.org/10.1016/j.jhydrol.2011.02.020.

Christensen, J., E. Kjellstrm, F. Giorgi, G. Lenderink, and R. M (2010), Weight assignment in regional climate models, *Climate Research*, *44*, 179–194, doi:10.3354/cr00916.

Christensen, J. H., and B. Hewitson (2007), Regional Climate Projections, in *CLIMATE CHANGE 2007: THE PHYSICAL SCIENCE BASIS*, edited by Solomon, S and Qin, D and Manning, M and Marquis, M and Averyt, K and Tignor, MMB and Miller, HL and Chen, ZL, pp. 847–940, CAMBRIDGE UNIV PRESS, 32 AVENUE OF THE AMERICAS, NEW YORK, NY 10013-2473 USA.

Christoffersen, B. O., et al. (2014), Mechanisms of water supply and vegetation demand govern the seasonality and magnitude of evapotranspiration in amazonia and cerrado, *Agricultural and Forest Meteorology*, *191*, 33 – 50, doi:https://doi.org/10.1016/j.agrformet.2014.02.008.

Chylek, P., J. Li, M. K. Dubey, M. Wang, and G. Lesins (2011), Observed and model simulated 20th century arctic temperature variability: Canadian earth system model canesm2, *Atmospheric Chemistry and Physics Discussions*, *11*, 22,893–22,907, doi:10.5194/acpd-11-22893-2011.

Cleveland, W. S., and S. J. Devlin (1988), Locally weighted regression: An approach to regression analysis by local fitting, *Journal of the American Statistical Association*, *83*(403), 596–610, doi:10.1080/01621459.1988.10478639.

Clow, D. W. (2010), Changes in the timing of snowmelt and streamflow in colorado: A response to recent warming, *Journal of Climate*, *23*(9), 2293–2306, doi:10.1175/2009JCLI2951.1.

Collins, M. D., R. E. Chandler, P. M. Cox, J. M. Huthnance, J. Rougier, and D. B. Stephenson (2012), Quantifying future climate change.

Costa, M. H., M. C. Biajoli, L. Sanches, A. C. M. Malhado, L. R. Hutyra, H. R. da Rocha, R. G. Aguiar, and A. C. de Arajo (2010), Atmospheric versus vegetation controls of amazonian tropical rain forest evapotranspiration: Are the wet and seasonally dry rain forests any different?, *Journal of Geophysical Research: Biogeosciences*, *115*(G4), doi:10.1029/2009JG001179.

Cunderlik, J. M., and T. B. Ouarda (2009), Trends in the timing and magnitude of floods in canada, *Journal of Hydrology*, *375*(3), 471 – 480, doi:https://doi.org/10.1016/j.jhydrol.2009.06.050.

Dai, A. (2013), Increasing drought under global warming in observations and models, *Nature Climate Change*, *3*(1), 52–58, doi:10.1038/nclimate1633.

Davidson, E. A., et al. (2012), The amazon basin in transition, *Nature*, *481*(7381), 321–328, doi:10.1038/nature10717.

de Abreu Sá, L. D., Y. Viswanadham, and A. O. Manzi (1988), Energy flux partitioning over the amazon forest, *Theoretical and Applied Climatology*, *39*(1), 1–16, doi:10.1007/BF00867653.

De Groeve, T., et al. (2015), Joining forces in a global flood partnership, *Bulletin of the American Meteorological Society*, *96*(5), ES97–ES100, doi:10.1175/BAMS-D-14-00147.1.

Dewar, R. C. (2005), Maximum entropy production and the fluctuation theorem, *Journal of Physics A: Mathematical and General*, *38*(21), L371–L381, doi:10.1088/0305-4470/38/21/l01.

Diaz-Nieto, J., and R. L. Wilby (2005), A comparison of statistical downscaling and climate change factor methods: impacts on low flows in the river thames, united kingdom, *Climatic Change*, *69*(2), 245–268, doi:10.1007/s10584-005-1157-6.

Dickinson, R. E., K. W. Oleson, G. Bonan, F. Hoffman, P. Thornton, M. Vertenstein, Z.-L. Yang, and X. Zeng (2006), The community land model and its climate statistics as a component of the community climate system model, *Journal of Climate*, *19*(11), 2302–2324, doi:10.1175/JCLI3742.1.

Dominguez, F., and P. Kumar (2008), Precipitation recycling variability and ecoclimatological stabilitya study using narr data. part i: Central u.s. plains ecoregion, *Journal of Climate*, *21*(20), 5165–5186, doi:10.1175/2008JCLI1756.1.

Doocy, S., A. Daniels, S. Murray, and T. D. Kirsch (2013), The human impact of floods: a historical review of events 1980-2009 and systematic literature review, *PLoS currents*, *5*, doi:10.1371/currents.dis.f4deb457904936b07c09daa98ee8171a.

d'Orgeville, M., W. R. Peltier, A. R. Erler, and J. Gula (2014), Climate change impacts on great lakes basin precipitation extremes, *Journal of Geophysical Research: Atmospheres*, *119*(18), 10,799–10,812, doi:10.1002/2014JD021855.

Dufresne, J.-L., et al. (2013), Climate change projections using the ipsl-cm5 earth system model: from cmip3 to cmip5, *Climate Dynamics*, *40*(9), 2123–2165, doi:10.1007/s00382-012-1636-1.

Dunne, J. P., et al. (2012), Gfdls esm2 global coupled climatecarbon earth system models. part i: Physical formulation and baseline simulation characteristics, *Journal of Climate*, *25*(19), 6646–6665, doi:10.1175/JCLI-D-11-00560.1.

Dwelle, M. C. (2018), Addressing variability in hydrologic systems using efficient uncertainty quantification, Ph.D. thesis, University of Michigan.

Dwelle, M. C., J. Kim, K. Sargsyan, and V. Y. Ivanov (2019), Streamflow, stomata, and soil pits: Sources of inference for complex models with fast, robust uncertainty quantification, *Advances in Water Resources*, *125*, 13 – 31, doi:https://doi.org/10.1016/j.advwatres.2019.01.002.

Emanuel, K. (2017), Assessing the present and future probability of hurricane harvey's rainfall, *Proceedings of the National Academy of Sciences*, *114*(48), 12,681–12,684, doi:10.1073/pnas.1716222114.

Farouki, O., C. R. Research, and E. L. (U.S.) (1981), *Thermal Properties of Soils*, CRREL monograph, U.S. Army Corps of Engineers, Cold Regions Research and Engineering Laboratory.

Fatichi, S., V. Y. Ivanov, and E. Caporali (2011), Simulation of future climate scenarios with a weather generator, *Advances in Water Resources*, *34*(4), 448 – 467, doi:https://doi.org/10.1016/j.advwatres.2010.12.013.

Fatichi, S., V. Y. Ivanov, and E. Caporali (2013), Assessment of a stochastic downscaling methodology in generating an ensemble of hourly future climate time series, *Climate Dynamics*, *40*(7), 1841–1861, doi:10.1007/s00382-012-1627-2.

Fatichi, S., V. Y. Ivanov, A. Paschalis, N. Peleg, P. Molnar, S. Rimkus, J. Kim, P. Burlando, and E. Caporali (2016), Uncertainty partition challenges the predictability of vital details of climate change, *Earth's Future*, *4*(5), 240–251, doi:10.1002/2015EF000336.

Fekete, B. M., C. J. Vrsmarty, and W. Grabs (2002), High-resolution fields of global runoff combining observed river discharge and simulated water balances, *Global Biogeochemical Cycles*, *16*(3), 15–1–15–10, doi:10.1029/1999GB001254.

Fischer, E. M., and R. Knutti (2016), Observed heavy precipitation increase confirms theory and early models, *Nature Climate Change*, *6*(11), 986–991, doi:10.1038/nclimate3110.

Fisher, J., et al. (2009), The land-atmosphere water flux in the tropics, *Global Change Biology*, *15*(11), 2694–2714, doi:10.1111/j.1365-2486.2008.01813.x.

Fowler, H. J., and M. Ekstrm (2009), Multi-model ensemble estimates of climate change impacts on uk seasonal precipitation extremes, *International Journal of Climatology*, *29*(3), 385–416, doi:10.1002/joc.1827.

Fowler, H. J., S. Blenkinsop, and C. Tebaldi (2007), Linking climate change modelling to impacts studies: recent advances in downscaling techniques for hydrological modelling, *International Journal of Climatology*, *27*(12), 1547–1578, doi:10.1002/joc.1556.

Gallego-Elvira, B., C. M. Taylor, P. P. Harris, and D. Ghent (2019), Evaluation of regional-scale soil moisture-surface flux dynamics in earth system models based on satellite observations of land surface temperature, *Geophysical Research Letters*, *46*(10), 5480–5488, doi:10.1029/2019GL082962.

Gent, P. R., et al. (2011), The community climate system model version 4, *Journal of Climate*, *24*(19), 4973–4991, doi:10.1175/2011JCLI4083.1.

Gerken, T., et al. (2018), Investigating the mechanisms responsible for the lack of surface energy balance closure in a central amazonian tropical rainforest, *Agricultural and Forest Meteorology*, *255*, 92 – 103, doi:https://doi.org/10.1016/j.agrformet.2017.03.023, honoring W.J. Massman's Discoveries: Bringing Physics to Agriculture.

Giorgetta, M. A., et al. (2013), Climate and carbon cycle changes from 1850 to 2100 in mpi-esm simulations for the coupled model intercomparison project phase 5, *Journal of Advances in Modeling Earth Systems*, *5*(3), 572–597, doi:10.1002/jame.20038.

Giorgi, F., and L. O. Mearns (2002), Calculation of average, uncertainty range, and reliability of regional climate changes from aogcm simulations via the reliability ensemble averaging (rea) method, *Journal of Climate*, *15*(10), 1141–1158, doi:10.1175/1520-0442(2002)015⟨1141:COAURA⟩2.0.CO;2.

Glenis, V., A. S. McGough, V. Kutija, C. Kilsby, and S. Woodman (2013), Flood modelling for cities using cloud computing, *Journal of Cloud Computing: Advances, Systems and Applications*, *2*(1), 7, doi:10.1186/2192-113X-2-7.

Gneiting, T., A. E. Raftery, A. H. Westveld, and T. Goldman (2005), Calibrated probabilistic forecasting using ensemble model output statistics and minimum crps estimation, *Monthly Weather Review*, *133*(5), 1098–1118, doi:10.1175/MWR2904.1.

G.R.Brakenridge (), Global active archive of large flood events, dartmouth Flood Observatory, University of Colorado, `http://floodobservatory.colorado.edu/Archives/index.html`.

Greve, P., L. Gudmundsson, and S. I. Seneviratne (2018), Regional scaling of annual mean precipitation and water availability with global temperature change, *Earth System Dynamics*, *9*(1), 227–240, doi:10.5194/esd-9-227-2018.

Gudmundsson, L., M. Leonard, H. X. Do, S. Westra, and S. I. Seneviratne (2019), Observed trends in global indicators of mean and extreme streamflow, *Geophysical Research Letters*, *46*(2), 756–766, doi:10.1029/2018GL079725.

Hajji, I., D. F. Nadeau, B. Music, F. Anctil, and J. Wang (2018), Application of the maximum entropy production model of evapotranspiration over partially vegetated water-limited land surfaces, *Journal of Hydrometeorology*, *19*(6), 989–1005, doi:10.1175/JHM-D-17-0133.1.

Hall, J. W., D. Grey, D. Garrick, F. Fung, C. Brown, S. J. Dadson, and C. W. Sadoff (2014), Coping with the curse of freshwater variability, *Science*, *346*(6208), 429–430, doi:10.1126/science.1257890.

Hanson, R. T., L. E. Flint, A. L. Flint, M. D. Dettinger, C. C. Faunt, D. Cayan, and W. Schmid (2012), A method for physically based model analysis of conjunctive

use in response to potential climate changes, *Water Resources Research*, *48*(6), doi:10.1029/2011WR010774.

Hasler, N., and R. Avissar (2007), What controls evapotranspiration in the amazon basin?, *Journal of Hydrometeorology*, *8*(3), 380–395, doi:10.1175/JHM587.1.

Haughton, N., G. Abramowitz, A. Pitman, and S. J. Phipps (2014), On the generation of climate model ensembles, *Climate Dynamics*, *43*(7), 2297–2308, doi:10.1007/s00382-014-2054-3.

Haughton, N., G. Abramowitz, A. Pitman, and S. J. Phipps (2015), Weighting climate model ensembles for mean and variance estimates, *Climate Dynamics*, *45*(11), 3169–3181, doi:10.1007/s00382-015-2531-3.

Hawkins, E., and R. Sutton (2009), The potential to narrow uncertainty in regional climate predictions, *Bulletin of the American Meteorological Society*, *90*(8), 1095–1108, doi:10.1175/2009BAMS2607.1.

Hay, L. E., R. L. Wilby, and G. H. Leavesley (2000), A comparison of delta change and downscaled gcm scenarios for three mountainous basins in the united states1, *JAWRA Journal of the American Water Resources Association*, *36*(2), 387–397, doi:10.1111/j.1752-1688.2000.tb04276.x.

Hilker, T., A. I. Lyapustin, F. G. Hall, R. Myneni, Y. Knyazikhin, Y. Wang, C. J. Tucker, and P. J. Sellers (2015), On the measurability of change in amazon vegetation from modis, *Remote Sensing of Environment*, *166*, 233 – 242, doi:https://doi.org/10.1016/j.rse.2015.05.020.

Hirabayashi, Y., R. Mahendran, S. Koirala, L. Konoshima, D. Yamazaki, S. Watanabe, H. Kim, and S. Kanae (2013), Global flood risk under climate change, *Nature Climate Change*, *3*(9), 816–821, doi:10.1038/nclimate1911.

Hirsch, R. M., and S. A. Archfield (2015), Not higher but more often, *Nature Climate Change*, *5*(3), 198–199, doi:10.1038/nclimate2551.

Hirsch, R. M., and K. R. Ryberg (2012), Has the magnitude of floods across the usa changed with global co2 levels?, *Hydrological Sciences Journal*, *57*(1), 1–9, doi:10.1080/02626667.2011.621895.

Hodgkins, G., R. Dudley, and T. Huntington (2003), Changes in the timing of high river flows in new england over the 20th century, *Journal of Hydrology*, *278*(1), 244 – 252, doi:https://doi.org/10.1016/S0022-1694(03)00155-0.

Hu, G., L. Jia, and M. Menenti (2015), Comparison of mod16 and lsa-saf msg evapotranspiration products over europe for 2011, *Remote Sensing of Environment*, *156*, 510 – 526, doi:https://doi.org/10.1016/j.rse.2014.10.017.

Hu, S., C. Zhao, J. Li, F. Wang, and Y. Chen (2014), Discussion and reassessment of the method used for accepting or rejecting data observed by a bowen ratio system, *Hydrological Processes*, *28*(15), 4506–4510, doi:10.1002/hyp.9962.

Huang, S.-Y., Y. Deng, and J. Wang (2017), Revisiting the global surface energy budgets with maximum-entropy-production model of surface heat fluxes, *Climate Dynamics*, *49*(5), 1531–1545, doi:10.1007/s00382-016-3395-x.

Huffman, G. J., D. T. Bolvin, E. J. Nelkin, D. B. Wolff, R. F. Adler, G. Gu, Y. Hong, K. P. Bowman, and E. F. Stocker (2007), The trmm multisatellite precipitation analysis (tmpa): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales, *Journal of Hydrometeorology*, *8*(1), 38–55, doi:10.1175/JHM560.1.

Hutyra, L. R., J. W. Munger, S. R. Saleska, E. Gottlieb, B. C. Daube, A. L. Dunn, D. F. Amaral, P. B. de Camargo, and S. C. Wofsy (2007), Seasonal controls on the exchange of carbon and water in an amazonian rain forest, *Journal of Geophysical Research: Biogeosciences*, *112*(G3), doi:10.1029/2006JG000365.

IPCC, 2018 (2018), Global warming of 1.5c. an ipcc special report on the impacts of global warming of 1.5c above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty [v. masson-delmotte, p. zhai, h. o. prtner, d. roberts, j. skea, p.r. shukla, a. pirani, w. moufouma-okia, c. pan, r. pidcock, s. connors, j. b. r. matthews, y. chen, x. zhou, m. i. gomis, e. lonnoy, t. maycock, m. tignor, t. waterfield (eds.)]., *Tech. rep.*

Ivancic, T. J., and S. B. Shaw (2015), Examining why trends in very heavy precipitation should not be mistaken for trends in very high river discharge, *Climatic Change*, *133*(4), 681–693, doi:10.1007/s10584-015-1476-1.

Ivanov, V. Y., E. R. Vivoni, R. L. Bras, and D. Entekhabi (2004), Catchment hydrologic response with a fully distributed triangulated irregular network model, *Water Resources Research*, *40*(11), doi:10.1029/2004WR003218.

Ivanov, V. Y., R. L. Bras, and D. C. Curtis (2007), A weather generator for hydrological, ecological, and agricultural applications, *Water Resources Research*, *43*(10), doi:10.1029/2006WR005364.

Ivanov, V. Y., R. L. Bras, and E. R. Vivoni (2008), Vegetation-hydrology dynamics in complex terrain of semiarid areas: 1. a mechanistic approach to modeling dynamic feedbacks, *Water Resources Research*, *44*(3), doi:10.1029/2006WR005588.

Ivanov, V. Y., S. Fatichi, G. D. Jenerette, J. F. Espeleta, P. A. Troch, and T. E. Huxman (2010), Hysteresis of soil moisture spatial heterogeneity and the homogenizing effect of vegetation, *Water Resources Research*, *46*(9), doi:10.1029/2009WR008611.

Ivanov, V. Y., L. R. Hutyra, S. C. Wofsy, J. W. Munger, S. R. Saleska, R. C. de Oliveira Jr., and P. B. de Camargo (2012), Root niche separation can explain

avoidance of seasonal drought stress and vulnerability of overstory trees to extended drought in a mature amazonian forest, *Water Resources Research*, *48*(12), doi:10.1029/2012WR011972.

Jacob, D., et al. (2014), Euro-cordex: new high-resolution climate change projections for european impact research, *Regional Environmental Change*, *14*(2), 563–578, doi:10.1007/s10113-013-0499-2.

Jakob Themel, M., A. Gobiet, and A. Leuprecht (2011), Empirical-statistical downscaling and error correction of daily precipitation from regional climate models, *International Journal of Climatology*, *31*(10), 1530–1544, doi:10.1002/joc.2168.

Jarvis, P. G., J. L. Monteith, and P. E. Weatherley (1976), The interpretation of the variations in leaf water potential and stomatal conductance found in canopies in the field, *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, *273*(927), 593–610, doi:10.1098/rstb.1976.0035.

Jasinski, M. F., et al. (2019), Nca-ldas: Overview and analysis of hydrologic trends for the national climate assessment, *Journal of Hydrometeorology*, *20*(8), 1595–1617, doi:10.1175/JHM-D-17-0234.1.

Jaynes, E. T. (2003), *Probability Theory: The Logic of Science*, Cambridge University Press, doi:10.1017/CBO9780511790423.

Jeffrey, S., L. D. Rotstayn, M. Collier, S. M. Dravitzki, C. Hamalainen, C. Moeseneder, K. Wong, and J. Syktus (2013), Australia s cmip 5 submission using the csiro-mk 3 . 6 model.

Jiménez-Muñoz, J. C., C. Mattar, J. Barichivich, A. Santamaría-Artigas, K. Takahashi, Y. Malhi, J. A. Sobrino, and G. v. d. Schrier (2016), Record-breaking warming and extreme drought in the amazon rainforest during the course of el niño 2015-2016, *Scientific Reports*, *6*, 33,130 EP –, article.

Jin, Y., J. T. Randerson, and M. L. Goulden (2011), Continental-scale net radiation and evapotranspiration estimated using modis satellite observations, *Remote Sensing of Environment*, *115*(9), 2302 – 2319, doi:https://doi.org/10.1016/j.rse.2011.04.031.

Jones, C. D., et al. (2011), The hadgem2-es implementation of cmip5 centennial simulations, *Geoscientific Model Development*, *4*(3), 543–570, doi:10.5194/gmd-4-543-2011.

Jurez, R. I. N., M. G. Hodnett, R. Fu, M. L. Goulden, and C. von Randow (2007), Control of dry season evapotranspiration over the amazonian forest as inferred from observations at a southern amazon forest site, *Journal of Climate*, *20*(12), 2827–2839, doi:10.1175/JCLI4184.1.

Kang, E. L., N. Cressie, and S. R. Sain (2012), Combining outputs from the north american regional climate change assessment program by using a bayesian hierarchical model, *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *61*(2), 291–313, doi:10.1111/j.1467-9876.2011.01010.x.

Karam, H. N., and R. L. Bras (2008), Climatological basin-scale amazonian evapotranspiration estimated through a water budget analysis, *Journal of Hydrometeorology*, *9*(5), 1048–1060, doi:10.1175/2008JHM888.1.

Keller, M., et al. (2004), Ecological research in the large-scale biosphere atmosphere experiment in amazonia: Early results, *Ecological Applications*, *14*(sp4), 3–16, doi:10.1890/03-6003.

Kennedy, M. C., and A. O'Hagan (2000), Predicting the output from a complex computer code when fast approximations are available, *Biometrika*, *87*(1), 1–13.

Kharin, V. V., F. W. Zwiers, X. Zhang, and M. Wehner (2013), Changes in temperature and precipitation extremes in the cmip5 ensemble, *Climatic Change*, *119*(2), 345–357, doi:10.1007/s10584-013-0705-8.

Kiang, J. E., et al. (2018), A comparison of methods for streamflow uncertainty estimation, *Water Resources Research*, *54*(10), 7149–7176, doi:10.1029/2018WR022708.

Kilsby, C., P. Jones, A. Burton, A. Ford, H. Fowler, C. Harpham, P. James, A. Smith, and R. Wilby (2007), A daily weather generator for use in climate change studies, *Environmental Modelling Software*, *22*(12), 1705 – 1719, doi:https://doi.org/10.1016/j.envsoft.2007.02.005.

Kim, J., and V. Y. Ivanov (2015), A holistic, multi-scale dynamic downscaling framework for climate impact assessments and challenges of addressing finer-scale watershed dynamics, *Journal of Hydrology*, *522*, 645 – 660, doi:https://doi.org/10.1016/j.jhydrol.2015.01.025.

Kim, J., A. Warnock, V. Y. Ivanov, and N. D. Katopodes (2012a), Coupled modeling of hydrologic and hydrodynamic processes including overland and channel flow, *Advances in Water Resources*, *37*, 104 – 126, doi:https://doi.org/10.1016/j.advwatres.2011.11.009.

Kim, J., V. Y. Ivanov, and S. Fatichi (2016), Climate change and uncertainty assessment over a hydroclimatic transect of michigan, *Stochastic Environmental Research and Risk Assessment*, *30*(3), 923–944, doi:10.1007/s00477-015-1097-2.

Kim, Y., R. G. Knox, M. Longo, D. Medvigy, L. R. Hutyra, E. H. Pyle, S. C. Wofsy, R. L. Bras, and P. R. Moorcroft (2012b), Seasonal carbon dynamics and water fluxes in an amazon rainforest, *Global Change Biology*, *18*(4), 1322–1334, doi:10.1111/j.1365-2486.2011.02629.x.

Knutti, R. (2010), The end of model democracy?, *Climatic Change*, *102*(3), 395–404, doi:10.1007/s10584-010-9800-2.

Knutti, R., and J. Sedláček (2013), Robustness and uncertainties in the new cmip5 climate model projections, *Nature Climate Change*, *3*(4), 369–373, doi:10.1038/nclimate1716.

Knutti, R., J. Sedlek, B. M. Sanderson, R. Lorenz, E. M. Fischer, and V. Eyring (2017), A climate model projection weighting scheme accounting for performance and interdependence, *Geophysical Research Letters*, *44*(4), 1909–1918, doi:10.1002/2016GL072012.

Koster, R. D., et al. (2004), Regions of strong coupling between soil moisture and precipitation, *Science*, *305*(5687), 1138–1140, doi:10.1126/science.1100217.

Leith, N. A., and R. E. Chandler (2010), A framework for interpreting climate model outputs, *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, *59*(2), 279–296.

Leuning, R., E. van Gorsel, W. J. Massman, and P. R. Isaac (2012), Reflections on the surface energy imbalance problem, *Agricultural and Forest Meteorology*, *156*, 65 – 74, doi:https://doi.org/10.1016/j.agrformet.2011.12.002.

Lewis, S. L., P. M. Brando, O. L. Phillips, G. M. F. van der Heijden, and D. Nepstad (2011), The 2010 amazon drought, *Science*, *331*(6017), 554–554, doi:10.1126/science.1200807.

Li, D., M. L. Wrzesien, M. Durand, J. Adam, and D. P. Lettenmaier (2017), How much runoff originates as snow in the western united states, and how will that change in the future?, *Geophysical Research Letters*, *44*(12), 6163–6172, doi:10.1002/2017GL073551.

Li, L., et al. (2013), The flexible global ocean-atmosphere-land system model, grid-point version 2: Fgoals-g2, *Advances in Atmospheric Sciences*, *30*(3), 543–560, doi:10.1007/s00376-012-2140-6.

Li, R., C. Wang, and D. Wu (2018), Changes in precipitation recycling over arid regions in the northern hemisphere, *Theoretical and Applied Climatology*, *131*(1), 489–502, doi:10.1007/s00704-016-1978-4.

Liang, X., D. P. Lettenmaier, E. F. Wood, and S. J. Burges (1994), A simple hydrologically based model of land surface water and energy fluxes for general circulation models, *Journal of Geophysical Research: Atmospheres*, *99*(D7), 14,415–14,428, doi:10.1029/94JD00483.

Lin, G., and G. E. Karniadakis (2009), Sensitivity analysis and stochastic simulations of non-equilibrium plasma flow, *International Journal for Numerical Methods in Engineering*, *80*(67), 738–766, doi:10.1002/nme.2582.

Lindroth, A., M. Mölder, and F. Lagergren (2010), Heat storage in forest biomass improves energy balance closure, *Biogeosciences*, *7*(1), 301–313, doi:10.5194/bg-7-301-2010.

Lins, H. F., and J. R. Slack (2005), Seasonal and regional characteristics of u.s. streamflow trends in the united states from 1940 to 1999, *Physical Geography*, *26*(6), 489–501, doi:10.2747/0272-3646.26.6.489.

Livneh, B., E. A. Rosenberg, C. Lin, B. Nijssen, V. Mishra, K. M. Andreadis, E. P. Maurer, and D. P. Lettenmaier (2013), A long-term hydrologically based dataset of land surface fluxes and states for the conterminous united states: Update and extensions, *Journal of Climate*, *26*(23), 9384–9392, doi:10.1175/JCLI-D-12-00508.1.

Long, D., L. Longuevergne, and B. R. Scanlon (2014), Uncertainty in evapotranspiration from land surface modeling, remote sensing, and grace satellites, *Water Resources Research*, *50*(2), 1131–1151, doi:10.1002/2013WR014581.

Lopes, A. V., J. C. H. Chiang, S. A. Thompson, and J. A. Dracup (2016), Trend and uncertainty in spatial-temporal patterns of hydrological droughts in the amazon basin, *Geophysical Research Letters*, *43*(7), 3307–3316, doi:10.1002/2016GL067738.

Maeda, E. E., X. Ma, F. H. Wagner, H. Kim, T. Oki, D. Eamus, and A. Huete (2017), Evapotranspiration seasonality across the amazon basin, *Earth System Dynamics*, *8*(2), 439–454, doi:10.5194/esd-8-439-2017.

Mahlstein, I., R. W. Portmann, J. S. Daniel, S. Solomon, and R. Knutti (2012), Perceptible changes in regional precipitation in a future climate, *Geophysical Research Letters*, *39*(5), doi:10.1029/2011GL050738.

Malhi, Y. (2012), The productivity, metabolism and carbon cycle of tropical forest vegetation, *Journal of Ecology*, *100*(1), 65–75, doi:10.1111/j.1365-2745.2011.01916.x.

Mallakpour, I., and G. Villarini (2015), The changing nature of flooding across the centralunited states, *Nature Climate Change*, *5*(3), 250–254, doi:10.1038/nclimate2516.

Mallick, K., et al. (2016), Canopy-scale biophysical controls of transpiration and evaporation in the amazon basin, *Hydrology and Earth System Sciences*, *20*(10), 4237–4264, doi:10.5194/hess-20-4237-2016.

Mamalakis, A., A. Langousis, R. Deidda, and M. Marrocu (2017), A parametric approach for simultaneous bias correction and high-resolution downscaling of climate model rainfall, *Water Resources Research*, *53*(3), 2149–2170, doi:10.1002/2016WR019578.

Manning, L. J., J. W. Hall, H. J. Fowler, C. G. Kilsby, and C. Tebaldi (2009), Using probabilistic climate change information from a multimodel ensemble for water resources assessment, *Water Resources Research*, *45*(11), doi:10.1029/2007WR006674.

Manoli, G., V. Y. Ivanov, and S. Fatichi (2018), Dry-season greening and water stress in amazonia: The role of modeling leaf phenology, *Journal of Geophysical Research: Biogeosciences*, *123*(6), 1909–1926, doi:10.1029/2017JG004282.

Maraun, D., et al. (2010), Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user, *Reviews of Geophysics*, *48*(3), doi:10.1029/2009RG000314.

Marengo, J. A., C. A. Nobre, J. Tomasella, M. D. Oyama, G. Sampaio de Oliveira, R. de Oliveira, H. Camargo, L. M. Alves, and I. F. Brown (2008), The drought of amazonia in 2005, *Journal of Climate*, *21*(3), 495–516, doi:10.1175/2007JCLI1600.1.

McHugh, M. L. (2013), The chi-square test of independence, *Biochemia Medica*, pp. 143–149, doi:10.11613/bm.2013.018.

Mearns, L. O., et al. (2013), Climate change projections of the north american regional climate change assessment program (narccap), *Climatic Change*, *120*(4), 965–975, doi:10.1007/s10584-013-0831-3.

Meehl, G., C. Covey, B. McAvaney, M. Latif, and R. Stouffer (2005), Overview of the Coupled Model Intercomparison Project, *BULLETIN OF THE AMERICAN METEOROLOGICAL SOCIETY*, *86*(1), 89–93, doi:{10.1175/BAMS-86-1-89}.

Meehl, G. A., et al. (2013), Climate change projections in cesm1(cam5) compared to ccsm4, *Journal of Climate*, *26*(17), 6287–6308, doi:10.1175/JCLI-D-12-00572.1.

Mehran, A., A. AghaKouchak, and T. J. Phillips (2014), Evaluation of cmip5 continental precipitation simulations relative to satellite-based gauge-adjusted observations, *Journal of Geophysical Research: Atmospheres*, *119*(4), 1695–1707, doi:10.1002/2013JD021152.

Milly, P. C. D., R. T. Wetherald, K. A. Dunne, and T. L. Delworth (2002), Increasing risk of great floods in a changing climate, *Nature*, *415*(6871), 514–517, doi:10.1038/415514a.

Milly, P. C. D., K. A. Dunne, and A. V. Vecchia (2005), Global pattern of trends in streamflow and water availability in a changing climate, *Nature*, *438*(7066), 347–350, doi:10.1038/nature04312.

Miranda, J. D. M. S. B. d. C. A. R., Rodrigo de Queiroga;Galvxed;ncio (2017), Reliability of modis evapotranspiration products for heterogeneous dry forest: A study case of caatinga, *Advances in Meteorology*, *2017*, doi:10.1155/2017/9314801.

Montanari, A., and D. Koutsoyiannis (2012), A blueprint for process-based modeling of uncertain hydrological systems, *Water Resources Research*, *48*(9), doi:10.1029/2011WR011412.

Moore, C., and G. Fisch (1986), Estimating heat storage in amazonian tropical forest, *Agricultural and Forest Meteorology*, *38*(1), 147 – 168, doi:https://doi.org/10.1016/0168-1923(86)90055-9.

Morss, R. E., J. L. Demuth, and J. K. Lazo (2008), Communicating uncertainty in weather forecasts: A survey of the u.s. public, *Weather and Forecasting*, *23*(5), 974–991, doi:10.1175/2008WAF2007088.1.

Mu, Q., F. A. Heinsch, M. Zhao, and S. W. Running (2007), Development of a global evapotranspiration algorithm based on modis and global meteorology data, *Remote Sensing of Environment*, *111*(4), 519 – 536, doi:https://doi.org/10.1016/j.rse.2007.04.015.

Mu, Q., L. A. Jones, J. S. Kimball, K. C. McDonald, and S. W. Running (2009), Satellite assessment of land surface evapotranspiration for the pan-arctic domain, *Water Resources Research*, *45*(9), doi:10.1029/2008WR007189.

Mu, Q., M. Zhao, and S. W. Running (2011), Improvements to a modis global terrestrial evapotranspiration algorithm, *Remote Sensing of Environment*, *115*(8), 1781 – 1800, doi:https://doi.org/10.1016/j.rse.2011.02.019.

Mueller, B., and S. I. Seneviratne (2014), Systematic land climate and evapotranspiration biases in cmip5 simulations, *Geophysical Research Letters*, *41*(1), 128–134, doi:10.1002/2013GL058055.

Mueller, B., et al. (2011), Evaluation of global observations-based evapotranspiration datasets and ipcc ar4 simulations, *Geophysical Research Letters*, *38*(6), doi:10.1029/2010GL046230.

Mueller, B., et al. (2013), Benchmark products for land evapotranspiration: Landflux-eval multi-data set synthesis, *Hydrology and Earth System Sciences*, *17*(10), 3707–3720, doi:10.5194/hess-17-3707-2013.

Najm, H. N. (2009), Uncertainty quantification and polynomial chaos techniques in computational fluid dynamics, *Annual Review of Fluid Mechanics*, *41*(1), 35–52, doi:10.1146/annurev.fluid.010908.165248.

Neal, J. C., T. J. Fewtrell, P. D. Bates, and N. G. Wright (2010), A comparison of three parallelisation methods for 2d flood inundation models, *Environmental Modelling Software*, *25*(4), 398 – 411, doi:https://doi.org/10.1016/j.envsoft.2009.11.007.

Nearing, G. S., M. S. Moran, R. L. Scott, and G. Ponce-Campos (2012), Coupling diffusion and maximum entropy models to estimate thermal inertia, *Remote Sensing of Environment*, *119*, 222 – 231, doi:https://doi.org/10.1016/j.rse.2011.12.012.

Nepstad, D. C., et al. (1994), The role of deep roots in the hydrological and carbon cycles of amazonian forests and pastures, *Nature*, *372*(6507), 666–669, doi:10.1038/372666a0.

Novick, K. A., et al. (2016), The increasing importance of atmospheric demand for ecosystem water and carbon fluxes, *Nature Climate Change*, *6*, 1023 EP –.

Nunes, J., J. Seixas, and J. Keizer (2013), Modeling the response of within-storm runoff and erosion dynamics to climate change in two mediterranean watersheds: A multi-model, multi-scale approach to scenario design and analysis, *CATENA*, *102*, 27 – 39, doi:https://doi.org/10.1016/j.catena.2011.04.001, scales in Soil Erosion.

Oki, T., and S. Kanae (2006), Global hydrological cycles and world water resources, *Science*, *313*(5790), 1068–1072, doi:10.1126/science.1128845.

Olson, R., Y. Fan, and J. P. Evans (2016), A simple method for bayesian model averaging of regional climate model projections: Application to southeast australian temperatures, *Geophysical Research Letters*, *43*(14), 7661–7669, doi:10.1002/2016GL069704.

Onyutha, C., H. Tabari, A. Rutkowska, P. Nyeko-Ogiramoi, and P. Willems (2016), Comparison of different statistical downscaling methods for climate change rainfall projections over the lake victoria basin considering cmip3 and cmip5, *Journal of Hydro-environment Research*, *12*, 31 – 45, doi:https://doi.org/10.1016/j.jher.2016.03.001.

Pan, S., H. Tian, S. R. Dangal, Q. Yang, J. Yang, C. Lu, B. Tao, W. Ren, and Z. Ouyang (2015), Responses of global terrestrial evapotranspiration to climate change and increasing atmospheric co2 in the 21st century, *Earth's Future*, *3*(1), 15–35, doi:10.1002/2014EF000263.

Peel, M. C., B. L. Finlayson, and T. A. McMahon (2007), Updated world map of the kppen-geiger climate classification, *Hydrology and Earth System Sciences*, *11*(5), 1633–1644, doi:10.5194/hess-11-1633-2007.

Peleg, N., S. Fatichi, A. Paschalis, P. Molnar, and P. Burlando (2017), An advanced stochastic weather generator for simulating 2-d high-resolution climate variables, *Journal of Advances in Modeling Earth Systems*, *9*(3), 1595–1627, doi:10.1002/2016MS000854.

Piani, C., J. O. Haerter, and E. Coppola (2010), Statistical bias correction for daily precipitation in regional climate models over europe, *Theoretical and Applied Climatology*, *99*(1), 187–192, doi:10.1007/s00704-009-0134-9.

Raftery, A. E., T. Gneiting, F. Balabdaoui, and M. Polakowski (2005), Using bayesian model averaging to calibrate forecast ensembles, *Monthly Weather Review*, *133*(5), 1155–1174, doi:10.1175/MWR2906.1.

Rasmussen, C. E. (2004), *Gaussian Processes in Machine Learning*, pp. 63–71, Springer Berlin Heidelberg, Berlin, Heidelberg, doi:10.1007/978-3-540-28650-9_4.

Renard, B., D. Kavetski, G. Kuczera, M. Thyer, and S. W. Franks (2010), Understanding predictive uncertainty in hydrologic modeling: The challenge of identifying input and structural errors, *Water Resources Research*, *46*(5), doi:10.1029/2009WR008328.

Restrepo-Coupe, N., et al. (2013), What drives the seasonality of photosynthesis across the amazon basin? a cross-site analysis of eddy flux tower measurements from the brasil flux network, *Agricultural and Forest Meteorology*, *182-183*, 128 – 144, doi:https://doi.org/10.1016/j.agrformet.2013.04.031.

Rienecker, M. M., et al. (2011), Merra: Nasas modern-era retrospective analysis for research and applications, *Journal of Climate*, *24*(14), 3624–3648, doi:10.1175/JCLI-D-11-00015.1.

Ripley, B. D. (1996), *Pattern Recognition and Neural Networks*, Cambridge University Press, doi:10.1017/CBO9780511812651.

Roache, P. J. (2001), Code Verification by the Method of Manufactured Solutions , *Journal of Fluids Engineering*, *124*(1), 4–10, doi:10.1115/1.1436090.

RISNEN, J. (2007), How reliable are climate models?, *Tellus A*, *59*(1), 2–29, doi:10.1111/j.1600-0870.2006.00211.x.

Saatchi, S., S. Asefi-Najafabady, Y. Malhi, L. E. O. C. Aragão, L. O. Anderson, R. B. Myneni, and R. Nemani (2013), Persistent effects of a severe drought on amazonian forest canopy, *Proceedings of the National Academy of Sciences*, *110*(2), 565–570, doi:10.1073/pnas.1204651110.

Saatchi, S. S., R. A. Houghton, R. C. Dos Santos Alvala, J. V. Soares, and Y. Yu (2007), Distribution of aboveground live biomass in the amazon basin, *Global Change Biology*, *13*(4), 816–837, doi:10.1111/j.1365-2486.2007.01323.x.

Safeeq, M., and A. Fares (2012), Hydrologic response of a hawaiian watershed to future climate change scenarios, *Hydrological Processes*, *26*(18), 2745–2764, doi:10.1002/hyp.8328.

Saito, K., S. Ogawa, and K. Otowa (2001), Estimates of lai for forest management in okutama.

Sargsyan, K., C. Safta, H. N. Najm, B. J. Debusschere, D. Ricciuto, and P. Thornton (2014), Dimensionality reduction for complex models via bayesian compressive sensing, *International Journal for Uncertainty Quantification*, *4*(1), 63–93.

Schmidli, J., C. Frei, and P. L. Vidale (2006), Downscaling from gcm precipitation: a benchmark for dynamical and statistical downscaling methods, *International Journal of Climatology*, *26*(5), 679–689, doi:10.1002/joc.1287.

Schoof, J., and S. Pryor (2001), Downscaling temperature and precipitation: a comparison of regression-based methods and artificial neural networks, *International Journal of Climatology*, *21*(7), 773–790, doi:10.1002/joc.655.

Schubert, J. E., and B. F. Sanders (2012), Building treatments for urban flood inundation models and implications for predictive skill and modeling efficiency, *Advances in Water Resources*, *41*, 49 – 64, doi:https://doi.org/10.1016/j.advwatres.2012.02.012.

Schubert, J. E., B. F. Sanders, M. J. Smith, and N. G. Wright (2008), Unstructured mesh generation and landcover-based resistance for hydrodynamic modeling of urban flooding, *Advances in Water Resources*, *31*(12), 1603 – 1621, doi:https://doi.org/10.1016/j.advwatres.2008.07.012.

Scoccimarro, E., S. Gualdi, A. Bellucci, A. Sanna, P. Giuseppe Fogli, E. Manzini, M. Vichi, P. Oddo, and A. Navarra (2011), Effects of tropical cyclones on ocean heat transport in a high-resolution coupled general circulation model, *Journal of Climate*, *24*(16), 4368–4384, doi:10.1175/2011JCLI4104.1.

Semenov, M. A., and P. Stratonovitch (2010), Use of multi-model ensembles from global climate models for assessment of climate change impacts.

Shanafield, M., P. G. Cook, H. A. Gutirrez-Jurado, R. Faux, J. Cleverly, and D. Eamus (2015), Field comparison of methods for estimating groundwater discharge by evaporation and evapotranspiration in an arid-zone playa, *Journal of Hydrology*, *527*, 1073 – 1083, doi:https://doi.org/10.1016/j.jhydrol.2015.06.003.

Sharma, A., C. Wasko, and D. P. Lettenmaier (2018), If precipitation extremes are increasing, why aren't floods?, *Water Resources Research*, *54*(11), 8545–8551, doi:10.1029/2018WR023749.

Shewchuk, J. R. (1996), Triangle: Engineering a 2D Quality Mesh Generator and Delaunay Triangulator, in *Applied Computational Geometry: Towards Geometric Engineering*, *Lecture Notes in Computer Science*, vol. 1148, edited by M. C. Lin and D. Manocha, pp. 203–222, Springer-Verlag, from the First ACM Workshop on Applied Computational Geometry.

Shi, T.-T., D.-X. Guan, J.-B. Wu, A.-Z. Wang, C.-J. Jin, and S.-J. Han (2008), Comparison of methods for estimating evapotranspiration rate of dry forest canopy: Eddy covariance, bowen ratio energy balance, and penman-monteith equation, *Journal of Geophysical Research: Atmospheres*, *113*(D19), doi:10.1029/2008JD010174.

Shuttleworth, W. J. (1993), chap. Chapter 4: 'Evaporation' [in] Handbook of hydrology, pp. 4.1–4.53, McGraw-Hill, New York, b1007352.

Shuttleworth, W. J., and H. C. Pereira (1988), Evaporation from amazonian rainforest, *Proceedings of the Royal Society of London. Series B. Biological Sciences*, *233*(1272), 321–346, doi:10.1098/rspb.1988.0024.

Slater, L. J., and G. Villarini (2016), Recent trends in u.s. flood risk, *Geophysical Research Letters*, *43*(24), 12,428–12,436, doi:10.1002/2016GL071199.

Smith, J. A., M. L. Baeck, G. Villarini, D. B. Wright, and W. Krajewski (2013), Extreme flood response: The june 2008 flooding in iowa, *Journal of Hydrometeorology*, *14*(6), 1810–1825, doi:10.1175/JHM-D-12-0191.1.

Smith, R. L., C. Tebaldi, D. Nychka, and L. O. Mearns (2009), Bayesian modeling of uncertainty in ensembles of climate models, *Journal of the American Statistical Association*, *104*(485), 97–116, doi:10.1198/jasa.2009.0007.

Sobol, I. (2001), Global sensitivity indices for nonlinear mathematical models and their monte carlo estimates, *Mathematics and Computers in Simulation*, *55*(1), 271 – 280, doi:https://doi.org/10.1016/S0378-4754(00)00270-6, the Second IMACS Seminar on Monte Carlo Methods.

Sommer, R., T. D. A. S, K. Vielhauer, A. C. de Arajo, H. Flster, and P. L. Vlek (2002), Transpiration and canopy conductance of secondary vegetation in the eastern amazon, *Agricultural and Forest Meteorology*, *112*(2), 103 – 121, doi:https://doi.org/10.1016/S0168-1923(02)00044-8.

Spracklen, D. V., S. R. Arnold, and C. M. Taylor (2012), Observations of increased tropical rainfall preceded by air passage over forests, *Nature*, *489*(7415), 282–285, doi:10.1038/nature11390.

Taylor, C. M., R. A. M. de Jeu, F. Guichard, P. P. Harris, and W. A. Dorigo (2012a), Afternoon rain more likely over drier soils, *Nature*, *489*(7416), 423–426, doi:10.1038/nature11377.

Taylor, K. E., R. J. Stouffer, and G. A. Meehl (2012b), An overview of cmip5 and the experiment design, *Bulletin of the American Meteorological Society*, *93*(4), 485–498, doi:10.1175/BAMS-D-11-00094.1.

Tebaldi, C., and R. Knutti (2007), The use of the multi-model ensemble in probabilistic climate projections, *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *365*(1857), 2053–2075, doi:10.1098/rsta.2007.2076.

Tebaldi, C., R. L. Smith, D. Nychka, and L. O. Mearns (2005), Quantifying uncertainty in projections of regional climate change: A bayesian approach to the analysis of multimodel ensembles, *Journal of Climate*, *18*(10), 1524–1540, doi:10.1175/JCLI3363.1.

ter Steege, H., et al. (2013), Hyperdominance in the amazonian tree flora, *Science*, *342*(6156), doi:10.1126/science.1243092.

Trenberth, K. E., J. T. Fasullo, and J. Kiehl (2009), Earth's global energy budget, *Bulletin of the American Meteorological Society*, *90*(3), 311–324, doi:10.1175/2008BAMS2634.1.

Twine, T., W. Kustas, J. Norman, D. Cook, P. Houser, T. Meyers, J. Prueger, P. Starks, and M. Wesely (2000), Correcting eddy-covariance flux underestimates over a grassland, *Agricultural and Forest Meteorology*, *103*(3), 279 – 300, doi: https://doi.org/10.1016/S0168-1923(00)00123-4.

Tyralis, H., and D. Koutsoyiannis (2017), On the prediction of persistent processes using the output of deterministic models, *Hydrological Sciences Journal*, *62*(13), 2083–2102, doi:10.1080/02626667.2017.1361535.

van Oldenborgh, G. J., et al. (2017), Attribution of extreme rainfall from hurricane harvey, august 2017, *Environmental Research Letters*, *12*(12), 124,009, doi:10.1088/1748-9326/aa9ef2.

van Vuuren, D. P., et al. (2011), The representative concentration pathways: an overview, *Climatic Change*, *109*(1), 5, doi:10.1007/s10584-011-0148-z.

Velpuri, N., G. Senay, R. Singh, S. Bohms, and J. Verdin (2013), A comprehensive evaluation of two modis evapotranspiration products over the conterminous united states: Using point and gridded fluxnet and water balance et, *Remote Sensing of Environment*, *139*, 35 – 49, doi:https://doi.org/10.1016/j.rse.2013.07.013.

Villarini, G. (2016), On the seasonality of flooding across the continental united states, *Advances in Water Resources*, *87*, 80 – 91, doi:https://doi.org/10.1016/j.advwatres.2015.11.009.

Voldoire, A., et al. (2013), The cnrm-cm5.1 global climate model: description and basic evaluation, *Climate Dynamics*, *40*(9), 2091–2121, doi:10.1007/s00382-011-1259-y.

Volodin, E. M., N. A. Dianskii, and A. V. Gusev (2010), Simulating present-day climate with the inmcm4.0 coupled model of the atmospheric and oceanic general circulations, *Izvestiya, Atmospheric and Oceanic Physics*, *46*(4), 414–431, doi:10.1134/S000143381004002X.

Von Storch, H. (2000), Review of empirical downscaling techniques, *Regional climate development under global warming. General technical report no. 4. Conference proceedings regclim spring meeting Jevnaker, Torbjornrud, Norway, 2000.*

Wang, B., and H. Yang (2008), Hydrological issues in lateral boundary conditions for regional climate modeling: simulation of east asian summer monsoon in 1998, *Climate Dynamics*, *31*(4), 477–490, doi:10.1007/s00382-008-0385-7.

Wang, J., and R. L. Bras (2009), A model of surface heat fluxes based on the theory of maximum entropy production, *Water Resources Research*, *45*(11), doi:10.1029/2009WR007900.

Wang, J., and R. L. Bras (2010), An extremum solution of the moninobukhov similarity equations, *Journal of the Atmospheric Sciences*, *67*(2), 485–499, doi:10.1175/2009JAS3117.1.

Wang, J., and R. L. Bras (2011), A model of evapotranspiration based on the theory of maximum entropy production, *Water Resources Research*, *47*(3), doi:10.1029/2010WR009392.

Wang, J., R. L. Bras, G. Sivandran, and R. G. Knox (2010), A simple method for the estimation of thermal inertia, *Geophysical Research Letters*, *37*(5), doi:10.1029/2009GL041851.

Wang, J., R. L. Bras, V. Nieves, and Y. Deng (2014), A model of energy budgets over water, snow, and ice surfaces, *Journal of Geophysical Research: Atmospheres*, *119*(10), 6034–6051, doi:10.1002/2013JD021150.

Wang, S., et al. (2015), Comparing evapotranspiration from eddy covariance measurements, water budgets, remote sensing, and land surface models over canada, *Journal of Hydrometeorology*, *16*(4), 1540–1560, doi:10.1175/JHM-D-14-0189.1.

Wasko, C., and A. Sharma (2017), Global assessment of flood and storm extremes with increased temperatures, *Scientific Reports*, *7*(1), 7945, doi:10.1038/s41598-017-08481-1.

Watanabe, M., et al. (2010), Improved climate simulation by miroc5: Mean states, variability, and climate sensitivity, *Journal of Climate*, *23*(23), 6312–6335, doi:10.1175/2010JCLI3679.1.

Wei, T., et al. (2012), Developed and developing world responsibilities for historical climate change and co2 mitigation, *Proceedings of the National Academy of Sciences*, *109*(32), 12,911–12,915, doi:10.1073/pnas.1203282109.

Weigel, A. P., R. Knutti, M. A. Liniger, and C. Appenzeller (2010), Risks of model weighting in multimodel climate projections, *Journal of Climate*, *23*(15), 4175–4191, doi:10.1175/2010JCLI3594.1.

Werth, D., and R. Avissar (2004), The regional evapotranspiration of the amazon, *Journal of Hydrometeorology*, *5*(1), 100–109, doi:10.1175/1525-7541(2004)005⟨0100:TREOTA⟩2.0.CO;2.

Widmann, M., C. S. Bretherton, and E. P. Salath (2003), Statistical precipitation downscaling over the northwestern united states using numerically simulated precipitation as a predictor, *Journal of Climate*, *16*(5), 799–816, doi:10.1175/1520-0442(2003)016⟨0799:SPDOTN⟩2.0.CO;2.

Wielicki, B. A., B. R. Barkstrom, E. F. Harrison, R. B. Lee, G. L. Smith, and J. E. Cooper (1996), Clouds and the earth's radiant energy system (ceres): An earth observing system experiment, *Bulletin of the American Meteorological Society*, *77*(5), 853–868, doi:10.1175/1520-0477(1996)077⟨0853:CATERE⟩2.0.CO;2.

Wilby, R. L., T. M. L. Wigley, D. Conway, P. D. Jones, B. C. Hewitson, J. Main, and D. S. Wilks (1998), Statistical downscaling of general circulation model output:

A comparison of methods, *Water Resources Research*, *34*(11), 2995–3008, doi:10.1029/98WR02577.

Williamson, G. B., W. F. Laurance, A. A. Oliveira, P. Delamnica, C. Gascon, T. E. Lovejoy, and L. Pohl (2000), Amazonian tree mortality during the 1997 el nio drought, *Conservation Biology*, *14*(5), 1538–1542, doi:10.1046/j.1523-1739.2000.99298.x.

Wilson, K., et al. (2002), Energy balance closure at fluxnet sites, *Agricultural and Forest Meteorology*, *113*(1), 223 – 243, doi:https://doi.org/10.1016/S0168-1923(02)00109-0, fLUXNET 2000 Synthesis.

Winsemius, H. C., et al. (2016), Global drivers of future river flood risk, *Nature Climate Change*, *6*(4), 381–385, doi:10.1038/nclimate2893.

Wood, A. W., L. R. Leung, V. Sridhar, and D. P. Lettenmaier (2004), Hydrologic implications of dynamical and statistical approaches to downscaling climate model outputs, *Climatic Change*, *62*(1), 189–216, doi:10.1023/B:CLIM.0000013685.99609.9e.

Wright, J. S., R. Fu, J. R. Worden, S. Chakraborty, N. E. Clinton, C. Risi, Y. Sun, and L. Yin (2017), Rainforest-initiated wet season onset over the southern amazon, *Proceedings of the National Academy of Sciences*, *114*(32), 8481–8486, doi:10.1073/pnas.1621516114.

Wu, J., et al. (2016), Leaf development and demography explain photosynthetic seasonality in amazon evergreen forests, *Science*, *351*(6276), 972–976, doi:10.1126/science.aad5068.

Wu, T. (2012), A mass-flux cumulus parameterization scheme for large-scale models: description and test with observations, *Climate Dynamics*, *38*(3), 725–744, doi:10.1007/s00382-011-0995-3.

Wyżga, B., Z. W. Kundzewicz, V. Ruiz-Villanueva, and J. Zawiejska (2016), *Flood Generation Mechanisms and Changes in Principal Drivers*, pp. 55–75, Springer International Publishing, Cham, doi:10.1007/978-3-319-41923-7_4.

Xiu, D., and G. E. Karniadakis (2002), The wiener–askey polynomial chaos for stochastic differential equations, *SIAM Journal on Scientific Computing*, *24*(2), 619–644, doi:10.1137/S1064827501387826.

Xu, D., V. Y. Ivanov, J. Kim, and S. Fatichi (2018), On the use of observations in assessment of multi-model climate ensemble, *Stochastic Environmental Research and Risk Assessment*, doi:10.1007/s00477-018-1621-2.

Xu, D., E. Agee, J. Wang, and V. Y. Ivanov (2019), Estimation of evapotranspiration of amazon rainforest using the maximum entropy production method, *Geophysical Research Letters*, *46*(3), 1402–1412, doi:10.1029/2018GL080907.

Yan, H., N. Sun, M. Wigmosta, R. Skaggs, L. R. Leung, A. Coleman, and Z. Hou (2019), Observed spatiotemporal changes in the mechanisms of extreme water available for runoff in the western united states, *Geophysical Research Letters*, *46*(2), 767–775, doi:10.1029/2018GL080260.

Yang, H., and B. Wang (2011), Reducing biases in regional climate downscaling by applying bayesian model averaging on large-scale forcing, *Climate Dynamics*, *39*, 2523–2532.

Yang, H., B. Wang, and B. Wang (2012), Reduction of systematic biases in regional climate downscaling through ensemble forcing, *Climate Dynamics*, *38*(3), 655–665, doi:10.1007/s00382-011-1006-4.

Yang, H., et al. (2017), Regional patterns of future runoff changes from earth system models constrained by observation, *Geophysical Research Letters*, *44*(11), 5540–5549, doi:10.1002/2017GL073454.

Yang, J., Z.-H. Wang, and T. W. Lee (2013), Relative efficiency of surface energy partitioning over different land covers, *International Journal of Environment and Climate Change*, pp. 186–102, doi:https://doi.org/10.9734/BJECC/2013/3062.

Yao, Y., et al. (2016), Assessment and simulation of global terrestrial latent heat flux by synthesis of cmip5 climate models and surface eddy covariance observations, *Agricultural and Forest Meteorology*, *223*, 151 – 167, doi:https://doi.org/10.1016/j.agrformet.2016.03.016.

Ye, S., H.-Y. Li, L. R. Leung, J. Guo, Q. Ran, Y. Demissie, and M. Sivapalan (2017), Understanding flood seasonality and its temporal shifts within the contiguous united states, *Journal of Hydrometeorology*, *18*(7), 1997–2009, doi:10.1175/JHM-D-16-0207.1.

Yin, L., R. Fu, E. Shevliakova, and R. E. Dickinson (2013), How well can cmip5 simulate precipitation and its controlling processes over tropical south america?, *Climate Dynamics*, *41*(11-12), 3127–3143, doi:10.1007/s00382-012-1582-y.

Yukimoto, S., et al. (2012), A new global climate model of the meteorological research institute: Mri-cgcm3 mdash;model description and basic performancemdash;, *Journal of the Meteorological Society of Japan. Ser. II*, *90A*, 23–64, doi:10.2151/jmsj.2012-A02.

Zhang, Z., K. Batselier, H. Liu, L. Daniel, and N. Wong (2017), Tensor computation: A new framework for high-dimensional problems in eda, *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, *36*(4), 521536, doi:10.1109/tcad.2016.2618879.

Zorita, E., and H. von Storch (1999), The analog method as a simple statistical downscaling technique: Comparison with more complicated methods, *Journal of Climate*, *12*(8), 2474–2489, doi:10.1175/1520-0442(1999)012⟨2474:TAMAAS⟩2.0.CO;2.